

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2001, 20:28:49 ; Search time 2326.71 Seconds
(without alignments)
11494.229 Million cell updates/sec

Title: US-09-597-513-1
Perfect score: 1729
Sequence: 1 tccacttcgcgcatttgaa.....tgaataaagggtggaactc 1729

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pl:*
- 48: em_rod:*
- 49: em_sta:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_v1:*
- 53: gb_sta1:*
- 54: gb_sta2:*
- 55: gb_sta3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_v11:*
- 59: gb_v12:*
- 60: gb_v13:*
- 61: gb_v14:*
- 62: gb_v15:*
- 63: gb_v16:*
- 64: gb_v17:*
- 65: gb_v18:*
- 66: gb_v19:*
- 67: gb_v20:*
- 68: gb_v21:*
- 69: gb_v22:*
- 70: gb_v23:*
- 71: gb_v24:*
- 72: gb_v25:*
- 73: gb_v26:*
- 74: gb_v27:*
- 75: gb_v28:*
- 76: gb_v29:*
- 77: gb_v30:*
- 78: gb_v31:*
- 79: gb_v32:*
- 80: gb_v33:*
- 81: gb_v34:*
- 82: gb_v35:*
- 83: gb_v36:*
- 84: gb_v37:*
- 85: gb_v38:*
- 86: gb_v39:*
- 87: gb_v40:*
- 88: gb_v41:*
- 89: gb_v42:*
- 90: gb_v43:*
- 91: gb_v44:*
- 92: gb_v45:*
- 93: gb_v46:*
- 94: gb_v47:*
- 95: gb_v48:*
- 96: gb_v49:*
- 97: gb_v50:*
- 98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1729	100.0	1729	1	AF005221	AF005221 Pseudomon
2	1725.8	99.8	20052	16	AF232006	AF232006 Pseudomon
3	1725.8	99.8	52498	2	AF232004	AF232004 Pseudomon
4	741.2	42.9	1833	1	AF037983	AF037983 Pseudomon
5	692.2	40.0	710	3	PS016119	U16119 Pseudomonas
6	155.2	9.0	2708	3	U94513	U94513 Erwinia amy
7	155.2	9.0	9985	2	EDSPAB	Y13831 Erwinia amy
8	153.6	8.9	4291	1	AF083620	AF083620 Erwinia a

9	67.4	3.9	727	6	DROMINI	M62837 Drosophila
10	65.8	3.8	78172	12	ABO10070	ABO10070 Arabidops
11	64.4	3.7	47852	3	MTV023	AL020202 Mycobacte
12	62.2	3.6	40549	3	SC2G38	AL445503 Streptomy
13	57.6	3.3	47739	1	AF017113	AF017113 Bacillus
14	57.6	3.3	209510	2	BS0B0018	Z99121 Bacillus su
15	57	3.3	10732	10	E32986	E32986 Gene encodi
16	56.8	3.3	76299	65	AC019585	AC019585 Drosophi
17	56.8	3.3	172363	4	AC009388	AC009388 Drosophi
18	56.8	3.3	203557	5	AE003763	AE003763 Drosophi
19	56.6	3.3	1245	1	AB011839	AB011839 Bacillus
20	56.2	3.3	1800	15	VCPEHEROPH	Y07752 V. carteri m
21	56	3.2	3100	3	HS062676	U62676 Halobacteri
22	56	3.2	10784	1	AE005080	AE005080 Halobacte
23	55.4	3.2	54327	78	AC090435	AC090435 Chlamydom
24	54.4	3.1	67200	3	MTV017	AL021897 Mycobacte
25	53.8	3.1	13205	3	RP4TRBAO	M93696 Plasmid RP4
26	53.8	3.1	60099	2	BIACOMGEN	L27758 Birmingham
27	53.8	3.1	119461	30	AC009249	AC009249 Drosophi
28	53.6	3.1	7218	10	I66494	I66494 Sequence 14
29	53.6	3.1	209156	83	CNS01DM4	AL136332 Homo sapi
30	53.4	3.1	63528	78	AC091196	AC091196 Homo sapi
31	53.2	3.1	16150	3	MTV044	AL021939 Mycobacte
32	52.6	3.0	67200	3	MTV017	AL021897 Mycobacte
33	52.4	3.0	618	14	CRGCR1	AL17207 Chlamydomon
34	52.4	3.0	53662	3	MTV016	AL021841 Mycobacte
35	52.2	3.0	6217	8	AB025226	AB025226 Gallus ga
36	52.2	3.0	167390	86	AC007263	AC007263 Homo sapi
37	52	3.0	1457	93	HSBHLH	X99268 H. sapiens m
38	51.8	3.0	38000	3	MTCT06H11	Z85982 Mycobacteri
39	51.8	3.0	47852	3	MTV023	AL020202 Mycobacte
40	51.8	3.0	68848	3	MTV043	AL022004 Mycobacte
41	51.8	3.0	204711	64	AC015936	AC015936 Homo sapi
42	51.4	3.0	1490	94	MMU70643	UT0643 Mus musculu
43	51.4	3.0	1546	94	MMU70642	UT0642 Mus musculu
44	51.4	3.0	1546	94	MMU70645	UT0645 Mus musculu
45	51.4	3.0	1547	94	MMU70649	UT0649 Mus musculu

ALIGNMENTS

RESULT	1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
--------	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

[illegible]

Db 1661 CACGAGCTTGATCCAGACAACTAGCTTCAAAAAAGGGGTGACTC 1729

RESULT 2

ID AF232006 standard; DNA; PRO; 20052 BP.

XX AF232006;

XX AF232006.1

XX 24-MAY-2000 (Rel. 63, Created)

DT 24-MAY-2000 (Rel. 63, last updated, Version 1)

XX Pseudomonas syringae pv. tomato strain DC3000 AVIE (avIE), HrpW (hrpW), and

DE Gsta (gsta) genes, complete cds; and unknown genes.

XX

XX Pseudomonas syringae pv. tomato

XX Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas

XX Pseudomonas syringae.

XX

XX [1]

XX 1-20052

XX MEDLINE; 98422476.

XX Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y., Collmer A.;

XX "The Pseudomonas syringae pv. tomato HrpW protein has domains similar to

XX harpins and peptate lyases and can elicit the plant hypersensitive response

XX and bind to peptate";

XX J. Bacteriol. 180(19):5211-5217(1998).

XX

XX [2]

XX 1-20052

XX MEDLINE; 20243785.

XX PUBMED; 10781092.

XX Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L., Petnicki-Ocwieja T.,

XX van Dijk K., Collmer A.;

XX "The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic

XX structure composed of a cluster of type III secretion genes bounded by

XX exchangeable effector and conserved effector loci that contribute to

XX parasitic fitness and pathogenicity in plants";

XX Proc. Natl. Acad. Sci. U.S.A. 97(9):4856-4861(2000).

XX

XX [3]

XX 1-20052

XX Alfano J.R., Collmer A.;

XX Submitted (07-FEB-2000) to the EMBL/GenBank/DBJ databases.

XX Dept. Biol. Sci., UNLV, 1854 Maryland Parkway, Las Vegas, NV 89154, USA

XX

XX SPTREMBL; 066101; 066101.

XX SPTREMBL; 087264; 087264.

XX SPTREMBL; 09JP27; 09JP27.

XX SPTREMBL; 09JP28; 09JP28.

XX SPTREMBL; 09JP29; 09JP29.

XX SPTREMBL; 09JP30; 09JP30.

XX SPTREMBL; 09JP31; 09JP31.

XX SPTREMBL; 09JP32; 09JP32.

XX SPTREMBL; 09JP33; 09JP33.

XX SPTREMBL; 09JP34; 09JP34.

XX SPTREMBL; 09JP35; 09JP35.

XX SPTREMBL; 09JP36; 09JP36.

XX SPTREMBL; 09JP37; 09JP37.

XX SPTREMBL; 09JP38; 09JP38.

XX

XX Key Location/Qualifiers

XX

XX 1. 20052

XX /db_xref="taxon:323"

XX /note="conserved effector locus flanking the hrpW/hrc

XX cluster pathovar: tomato"

XX /organism="Pseudomonas syringae pv. tomato"

XX


```

|||||
Db 11330 CCGTGTGAGTGGGGGAGAGAAATACAGTCTTCTGCGGTGCTGACACTAGTCGC 11389
|||
Qy 181 aggcataaggcattcagttcctctgctgcttgggcatataaaaaaaggaaactttaaa 240
|||||
Db 11390 AGGATAGGATTCAGTCTCTGCTGCTGGTGGCATATATAAAAAAGAACTTTTAAAA 11449
|||
Qy 241 acagtgaaatgagatgctggcgaacgaaggaacgggtcgtcgtcgtcttggcaactactcg 300
|||||
Db 11450 ACAGTGCAATGAGATGCGGCGGCAAAACGGGAACCGGTGCGCGCTTTCGCACTCATTGC 11509
|||
Qy 301 agcaagctcaaccaccaacatccacatccatctgaaacggaacggaatacggccaacttgc 360
|||||
Db 11510 AGCAAGCTCAACCCCAACAAACATCCACATCCTATGAAACGGACAGCATAGCGCATTTGC 11569
|||
Qy 361 tctgtgtaaacctctggagctgctgctgcttccaatgtgccacttagcgaagtaacgcaat 420
|||||
Db 11570 TCTGTGTAAACCTTGAGACTGGCGGTGCTCAATTGCCACTTAGCGAGTAACGACAT 11629
|||
Qy 421 gagcatcgcatataacacccggccgcaacgaacgaacgaacgaacgaacgaacgaacgaac 480
|||||
Db 11630 GAGATCGGATCAACACCCGCGCGCAACAGACCAACGCGCATCTGATTTTTCGGCGCT 11689
|||
Qy 481 aagcggcaagatgctcctaacaacaacacgcttcggcggaacgaacgaacgaacgaacgaac 540
|||||
Db 11690 AAGCGGCAAGTCTCAACCAAAACAGCTTCGGGAGACAGACACTAGCAAGCATCGA 11749
|||
Qy 541 ccgagatgacgtctgtctggcagcgacacacagaagaagcgaactcggcgcgcga 600
|||||
Db 11750 CCGGAGTGCCTGCTTGTTCGGGAGCGACACAGAAAGAGTCAACTTTCGGACGCCGA 11809
|||
Qy 601 cagcacccgtcagaatctcggaaggaagcgaagcccaaggaaggaagcgaagcgaacatcgc 660
|||||
Db 11810 CAGCACCGTCAAGAAATCCGAGAGCGCACAAAGCCCAAGGACGCCAAGTCCAAACATCGC 11869
|||
Qy 661 taatgatcatgcatgtcatgtctgtctgcaagatctcaaccaaccacaataaaaa 720
|||||
Db 11870 TAAATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11929
|||
Qy 721 gcaagacacacacacacacacacacacacacacacacacacacacacacacacacacacac 780
|||||
Db 11930 GCGAGGACACCAATCAGGAAACAGCTGATACCGAGGCTCTTTCAGAAACAGCGCGGCT 11989
|||
Qy 781 cgttacaacgtcgccgatgctggggcgggcggttacaacggatgctgcaacggttggcgcg 840
|||||
Db 11990 CGGTACACCGTCCGCGATGACGGGGGCGGCGGTACACCGGATGCGCACAGTGGCGCGG 12049
|||
Qy 841 cgttatacgcgaacgcgaacgcgaacgcgttggcggttgcgttgcgttgcgttgcgttgcgtt 900
|||||
Db 12050 CGGTGATACGCCCAAGCGCAACAGCGGCGGTGCGGTGATGATGATGATGATGATGATGATGAT 12109
|||
Qy 901 tggcggaacggttggcgcggaacacacacacacacacacacacacacacacacacacacacac 960
|||||
Db 12110 TGGCGGACGAGGCTGGGCGGCGCACACCACTGCAACAGGTGGCGGCGGTGGCACACC 12169
|||
Qy 961 cactgcaaacgaggttggcggaaggttggcgtaaacacgcaataactccggaattggcaca 1020
|||||
Db 12170 CACTGCAACAGGCGGTGGGAGGCGGTGCGTAACACGCAATATCTCCGAGTTGGCGCAA 12229
|||
Qy 1021 ccctaacacgttaccaggttactgctcgttgcgttgcgaacacgcggttcttccgaagcaagc 1080
|||||
Db 12230 CCTTAACCGTACTCAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 12289
|||
Qy 1081 cggcaagaatcaatgtgtgtaagaacacacacacacacacacacacacacacacacacacacac 1140
|||||
Db 12290 CGGCAAGATCAATGTGTGTAAGAACACATCAAGGTGCGGCGTAATCTTTTGACGG 12349
|||
Qy 1141 ccacggcgcaacactcactcgcacacaatctatgggttaacggagacacagggcgaaataca 1200
|||||
Db 12350 CCACGGCGCAACTTCTACTGCTCCGACAAATCTATGGGTAAGGGAGACCGGCGGCAAAATCA 12409
|||
Qy 1201 gaagcccatgttcgaagctggtcgaagcgctgaagtaaaagtgaaccctgggtgagaa 1260
|||||

```

```

Db 12410 GAAGCCATGTTGAGCTGGCTGAGAGGCGCTACGTTGAAGATGTGAACCTGGGTGAGAA 12469
|||
Qy 1261 cgaagtcgatgtgcatccacgttgaagaccaaacaacgctcagaagtcatccattgacaacgt 1320
|||||
Db 12470 CGAGTGCATGGCATTCACAGCTGAAGGCCAAAACGCTCAGAGATGCACATTGACACAGT 12529
|||
Qy 1321 gcatgcccagaacgttgggtgaagacacctgattacggttaaaaggcggaaggcgacgagct 1380
|||||
Db 12530 GCATGCCCAACACGTCGTTGAAGACCTGATTAACGCTCAAAAGCGCAGGAGCGCGACGCT 12589
|||
Qy 1381 cactaatctgaacatacaagaacacgagtgccaaagtgctgcagaacgaaggttgcagct 1440
|||||
Db 12590 CACTAATCTGATCAATCAAGAACAGCATGCCCAAGGTGCGACAGCATGTTGTCAGCT 12649
|||
Qy 1441 caacgccaacactcacttgaanaatcgacaacttcaaaagccgacgatttcggcagatgct 1500
|||||
Db 12650 CAACGCCCAACATCACTTGAATAATCGACAACCTTCAAGCGCGAGATTTTCGACAGATGCT 12709
|||
Qy 1501 tgcgaccaacaggttggcggaacgagtttgatgacatgagcatgagctgaagcgatcgaaac 1560
|||||
Db 12710 TCGCACCAACGGTGGCAAGGAGTTTGTATGATGATGATGATGATGATGATGATGATGATGAT 12769
|||
Qy 1561 taaccaagcgaagcttgcacctgtgtgaagaacgcaacgagtgacatctgaagcttgcgaacgg 1620
|||||
Db 12770 TAACACGGCAAGTTCGCTCCTGTGTAAGAACGACAGCATGATGATGATGATGATGATGATGAT 12829
|||
Qy 1621 caaatcgcacatgacgcgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1680
|||||
Db 12830 CAACATGCGCATGACCGACGCTCAACACGCTTACGATTAACCAACGAGCATGACACCAACA 12889
|||
Qy 1681 caccgagcttgaatccagcgaagtgcttgaanaaagggtgagatc 1729
|||||
Db 12890 CACCGAGCTTTGATTCAGACAAAGTACCTTGAATAAAAGGGGTGAGACTC 12938
|||

```

RESULT 3
AF232004
LOCUS 52498 bp DNA BCT 05-MAR-2001
DEFINITION Pseudomonas syringae pv. tomato strain DC3000 Hrp pathogenicity island, complete sequence.
ACCESSION AF232004 AF061028 AF061029 AF232006 I41861
VERSION AF232004.3 GI:13325077
KEYWORDS
SOURCE Pseudomonas syringae pv. tomato.
ORGANISM Pseudomonas syringae pv. tomato Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 25494 to 29778) Preston, G., Huang, H.C., He, S.Y. and Collmer, A. The Hrp proteins of *Pseudomonas syringae* pvs. syringae, glycinea, and tomato are encoded by an operon containing *Yersinia* ysc homologs and elicit the hypersensitive response in tomato but not soybean.
JOURNAL Mol. Plant Microbe Interact. 8 (5), 717-732 (1995)
MEDLINE 96025089
PUBMED 7579616
REFERENCE 2 (bases 22134 to 25847; 29687 to 32670) Deng, W.L., Preston, G., Collmer, A., Chang, C.J. and Huang, H.C. Characterization of the hrpC and hrpR operons of *Pseudomonas syringae* pathovars syringae, tomato, and glycinea and analysis of the ability of hrpF, hrpG, hrpC, hrpR, and hrpV mutants to elicit the hypersensitive response and disease in plants.
JOURNAL J. Bacteriol. 180 (17), 4533-4531 (1998)
MEDLINE 98389667
PUBMED 9721291
REFERENCE 3 (bases 31672 to 51723) Charkowski, A.O., Alfano, J.R., Preston, G., Yuan, J., He, S.Y. and Collmer, A. The *Pseudomonas syringae* pv. tomato HrpW protein has domains similar to harpins and pectate lyases and can elicit the plant hypersensitive response and bind to pectate.
JOURNAL J. Bacteriol. 180 (19), 5211-5217 (1998)
MEDLINE 98422476

PUBMED 9748456
 REFERENCE 4 (bases 901 to 22404; 31672 to 51723)
 AUTHORS Alfano, J.R., Charkowski, A.O., Deng, W.L., Badel, J.L.,
 TITLE Penicki-Ocwieja, T., van Dijk, K. and Collmer, A.
 The pseudomonas syringae hrp pathogenicity island has a tripartite
 mosaic structure composed of a cluster of type III secretion genes
 bounded by exchangeable effector and conserved effector loci that
 contribute to parasitic fitness and pathogenicity in plants
 Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4856-4861 (2000)
 JOURNAL MEDLINE 20243785
 PUBMED 10781092
 REFERENCE 5 (bases 1 to 52498)
 AUTHORS Ramos, A.R., Rehm, A.H. and Collmer, A.R.
 TITLE Pseudomonas syringae pv. tomato DC3000 hrp1 through hrp11
 JOURNAL Unpublished
 REFERENCE 6 (bases 1 to 52498)
 AUTHORS Alfano, J.R. and Collmer, A.
 TITLE Direct Submission
 JOURNAL Submitted (07-FEB-2000) Dept. Biol. Sci., UNLV, 1854 Maryland
 Parkway, Las Vegas, NV 89154, USA
 REFERENCE 7 (bases 1 to 52498)
 AUTHORS Ramos, A.R., Rehm, A.H. and Collmer, A.R.
 TITLE Direct Submission
 JOURNAL Submitted (22-NOV-2000) Plant Pathology, Cornell University, 334
 Plant Sciences Bldg., Ithaca, NY 14850, USA
 REFERENCE 8 (bases 1 to 52498)
 AUTHORS Ramos, A.R., Rehm, A.H. and Collmer, A.R.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAR-2001) Plant Pathology, Cornell University, 334
 Plant Sciences Bldg., Ithaca, NY 14850, USA
 REMARK Sequence update by submitter
 COMMENT On Mar 14, 2001 this sequence version replaced gi:3228544 gi:790906
 gi:3228541 gi:8037790 gi:11276506.
 FEATURES
 SOURCE Location/Qualifiers
 1..52498
 /organism="Pseudomonas syringae pv. tomato"
 /strain="DC3000"
 /db_xref="taxon:323"
 /note="exchangeable effector locus flanking the hrp/hrc
 cluster
 pathovar: tomato"
 complement(908..1972)
 /gene="guea"
 complement(908..1972)
 /gene="guea"
 /codon_start=1
 /transl_table=11
 /product="Guea"
 /protein_id="AAF71483.1"
 /db_xref="GI:8037772"
 /translation="MRVADTFEELPDSLIARHPLAERRSRLLTLDDGTGALAHROPT
 DLIELHSGDLMVFNRTVITAPRLFGCAKSGKLEILVERVLDSHRYLAHRSKPK
 FGSSILIDGGGEAEHVARHDALEFLREFEVLPLIDRVGHMPLPYIDRDEGADRRK
 YQTVYAQRAGAVAAPITAGLHFDPLMEIAKGVETAFVTLHVGAGTFQPYRQVIED
 HHMSSEMLEVSQDVDAVAACRARGRYIAGTISVRSLESAARDGOLKPKSGPTDIF
 IYGPREFVVDALVTNFIPLPESTLIMLVASAGVPEFWAAVAAALIEHGYRFSYGDAM
 EITRNPAVTAQESAPEDHA"
 2090..2177
 /product="tRNA-Leu"
 2343..2753
 /note="ORF4"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAF71484.1"
 /db_xref="GI:8037773"
 /translation="MNKIVYKAYFKPIGEEVSVKVPGLGEIKKFGFGDKETMKETQW
 QQTGMSQDIDGERTLSDVDVAVALNADGEIQTQVLPILSGAVDYALKYREIRHNR
 TELSPGOSYQIIEGYSFTGEGVTLVAKKFSAS"
 complement(2908..3324)
 /note="ORF3"
 /codon_start=1

CDS
 /transl_table=11
 /product="unknown"
 /protein_id="AAF71485.1"
 /db_xref="GI:8037774"
 /translation="MRAYKNTAKIGFLLATITIGSLPFAVNCDDLDNDNSTGAT
 CGGNDKDLNDNVDAALFGGNDKMDNDHTDAFGGNDKDLNDNDHTDAFGGNDK
 LDNDNKTDAAFGGNDKDLNDNDNDNNGPSPAAKK"
 complement(3939..4664)
 /note="ORF2"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAF71486.1"
 /db_xref="GI:8037775"
 /translation="MOKTTLAALFAMLAGCGVSPAGSDIOGAQAEKTPYKLNLD
 AVTSKRLDAVLEARTKSYMKRGOLIDHGAFLGTPYRSMTIGSKANVPEQLYDPR
 GLDCEPATIDYEAFRRSTSQDPEYRNLYQVRYKSGDYDELKRFKFFPDMAVGTYPRA
 DDITQAISPGAVSVKRLNERAKGVYLPGLVVERSMYTIPLSLVDSQVYVSHRTGD
 YIGIYTPSRAGCDTRRFLYRDG"
 5571..6410
 /gene="tnpA"
 5571..6410
 /gene="tnpA"
 /codon_start=1
 /transl_table=11
 /product="tnpA"
 /protein_id="AAF71487.1"
 /db_xref="GI:8037776"
 /translation="MGLRPAYPLMAMRLHLMONWPDYSDPAMEALYETTLRQRA
 GLSLDRIDAEYTLILFRLLERLHLAGGILQVINGYLGRGLMLRQGVADATILHAP
 LSTNKKGKRDPEMHTQTKGNQYFGKGAHLGVASGLVHSLVGTANADAVTYQVQ
 LLHSEETVSGDAGYTGDKRAHQDQIMSIARSRYSKHEKSLIARVYKIEF
 TKAOLRAKVEHPLRVIKRQFGYTVRRFLAKNNAQAQATLFLSNLMWVRRLIAME
 VRL"
 complement(6656..8056)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAF71488.1"
 /db_xref="GI:8037777"
 /translation="MRPVGGPAPGYPTTYEAEPTLOAANCNDRASSQSSPAAVY
 APETPMGLDKRFPAGRYPDMKVENIRLKEQEPGKDKGKHRRRRPKPAAGSSVHH
 GQGSVASTSASQSKALQDPTNFKAIDLELRWESPYPVALSPSKAGSSQSLNAVY
 VSILOGKHALBEOLLEAGLADVYVSECDLHLINLTLEMDSGISGGLMAD
 SNDKILIAKARYDFDFAOKPLPLAPLTKKSDGVMEILDLGALVYGHNS
 TSSKRELIINMKSLKASGVITLPMHEHLCASHDKALNNYLSAPKSGMPARAKRYIDL
 OSQGHQAPBEELHTKYNFTTLVEAKHAGLRVSLDTYSTMAPEKAEIKRQANNTYA
 AEKIRLSKPEGKWVAFVGAHTATSCDGVPGIAELHGVRSIVIDDLGIKSRATVDINVK
 NYGKLNPDVRLSYKV"
 complement(8131..10473)
 /gene="hrpK"
 complement(8131..10473)
 /gene="hrpK"
 /codon_start=1
 /transl_table=11
 /product="hrpK"
 /protein_id="AAF71489.1"
 /db_xref="GI:8037778"
 /translation="MRISSEPFVIVNOPTPEELALAVESPLAKLPLPPVGGGQAGVQ
 FGDPAGNTOGAPGAEQTASSILALLQSGGAANANNVNOVQPOPCADSPAPAAFT
 EATAPAAVEQVAPVTWTDATDASNARADAAFLDENSESSPEALRWEPVVAHLP
 EERQAKKELNRPITAAAMMARDDGPDGAKMDFINANPALKTAVDVAOGGNADGKIT
 NKDLKTPAKNMERKADSADLDLAVYTRKDNPDPOSLEWYRSAPVEMRANPPLATAADP
 HHVADDAKTRVDSNDVADIEDIKLAGSNPGLSGALKOSCSYWSQPGTIGVDEAGMG
 RKAHSPDKRMEDAKNLSMTTKSAPTNNGQFASMLSDSATLNAVAGIDISKDGVF
 DKPSYSYGAQAAVAVKILQQTQDSVILAGSRNRTDKTEQGLNDRI SOLQADPDQATL
 NKSIPQEBRNLYRSDASLOKAVVEQTRNVSGGALQTDMDKADKAVKHNPNADYSL
 ISGSAOLOLOKIDPDSKVTYTNQVLENKPDLDQKITATSVYTFSEGGALKQLLGK
 KSDASOSILOADNOKATYDYLTPDPFENQENWNSLTGIDQSKGRGLBEKSNQ
 GGSMTLOLAEGYCGKALNSYGFASVSSILIAAGDLGATQATLYDSTRGAEIRKSG
 IDAGAKMGRASAGLGRLGOMTGRAVWAGATGATLAGAALGAALPYVGAVIDGA
 LAVGIGISMTIIDAVKHKRAQKAPDNDPVLIDQGIKAAH"

gene		complement(10560..11114)
CDS		/gene="hrpL" complement(10560..11114) /gene="hrpL" /note="alternative sigma factor" /codon_start=1 /transl_table=11 /product="HrpL" /protein_id="AAG3875.2" /db_xref="GI:13325078" /translation="MFQKIVILIDSTORPOSSAGIROMADQIOMLRAPFIOKRVNMPDNDVLIQCVELALRHEHKFRQHSKPOTMLIALNLINHPKRFQROSEMEDEVHSLTEGHGVSHOVESHROLARYIQIIDLPLSNMQRYLEVLEMDNTQETANLSGVPIGVRSLSGARFOLOKDPEFA"
gene		11358..12464
CDS		/gene="hrpJ" 11358..12464 /gene="hrpJ"
Query Match	99.8%; Score 1725.8; DB 2; Length 52498;	
Best Local Similarity	99.9%; Pred. No. 0;	
Matches 1727; Conservative	0; Mismatches 2; Indels 0; Gaps 0;	
Oy	1	tccacttcgcgtgattttgaatatggcacaatccataagaacgttcaagtgttgaaatcagg 60
Dd	42881	TCCACTTCGCCTGAATTTTAAATGGCAATTCATGTGAACCGTGCGTAATTCAGG 42940
Oy	61	ctgagtcgcagatttcgttgttaagggtgtgtgtactgtcatctgttgatcttcaaag 120
Dd	42941	CTGAGTCGCAGATTTCGTTGTGATTAAGGGTGTGTACTGTCTGTCATTGTGTCATTTCAAAG 43000
Oy	121	cctctgagtcggtgcggaagacaataccagcttccctgcgtgtgcgtgtgtgcacatgagtcgc 180
Dd	43001	CCTCTGAGTCGGGTGCGGACGAATACGATCCTGCTGGTGGGTGTGCACATGAGTGC 43060
Oy	181	aggcataggcaattcagttccttgcgttgtgtgtgtggcataaaaaaaggaactttaaa 240
Dd	43061	AGGCATAGGCATTTCA GTTCCTTGCGTGGTGCGCATATAAAAAAGAAC TTTTAAA 43120
Oy	241	aagatgcgatgagatgcgcggaacacgggaacaggttcgtcggtttccaactcttcg 300
Dd	43121	ACAGTGCATAGAGATGCCGGCAAAAACGGGAACGGTCCGTGCGCTTGTGCATCTTCG 43180
Oy	301	agcaagctcaaaccccaaatcacatccatactctaatacgaacgacgatalacgcccacttgc 360
Dd	43181	AGCAAGCTCAACCCCAAAACATCCACATCCCTATCTGAACGAGACGGATNAGGCCACTTGC 43240
Oy	361	tctgttaaacccttggagcttgcgttcgttccaatttgcaccttagcgaagtaacgacat 420
Dd	43241	TCTGTAAACCCTTGAGCTGGGTGGTGTGCATATCCCACTTAGGAGAGTAACGACGAT 43300
Oy	421	ggagctggcaataccccggcgcgcaacagacacacgcgcacatcgatttttcgggct 480
Dd	43301	GAGCATTGGCATCACACCCCGGCCCAACAGCCACACCGCATCTGATTTTTCGGCGCT 43360
Oy	481	aagcggcaagaagttcctcaaccaaaaacglttcggcgagagaagaacacttcagcaagcgatcga 540
Dd	43361	AAGCGCAAGAAGTCTCTCAACCAAAACGTTGGCGAGCAANAAC TCTAGCAAGCATGCA 43420
Oy	541	cccgggtgactgttcttcggcgaagcgaacacagaagaagaagtcacttcggaagcccca 600
Dd	43421	CCCGAGTGCACATGTTGTTCCGACAGGACAACAAGAAAGCGTCAACTTTCGGACGCCGA 43480
Oy	601	cagacccgtccagaatccgcaagacgacgaacgacgaacgaacagcagtcacagtcacacatgc 660
Dd	43481	CAGCACCGTCCAGNAATCCGCAAGACGCCAAGCAAGCCCAACGACACCGATCAACATGCG 43540
Oy	661	taaatgtacagtgcattgatatgtctgtctgtagatgtctcacaaatccaataaaaa 720
Dd	43541	TAAATGTGATCAATGATTCATGTCTGTTGTGAGATGTCACCAATCCCAATAAAAA 43600
Oy	721	gcaggaacacaaatcaggaacagcctgtatagcaggctcccttccagaaacagcgcgagct 780
Dd		

Db	43601	GCAGGACACACATATAGAGAAACACCTCGATATGACGAGCTCTTTCCAGAACACAGGGGGGT	43668
Qy	781	cggtacacccgttcgcgcgaatagcggggcgcggtgataccggatctgcgaagttgvcgvcg	840
Db	43661	CGGTACACCGTGTGGCCGATATAGCGGGGGCGCGGTACCCGATGCGACAGTGGCGGGCG	43720
Qy	841	cggtatatacgccaagcgcgaacagcggtgvcgvcggttgatatctccgaccgcgaacggcg	900
Db	43721	CGGTATATCGCCACACCGCAACGACGGGGGTGGCGCGGTGTATCTCCGACCGCAACAGGGCG	43780
Qy	901	tggcggaagcgggtggcgcgcgcgcacacccactgcaacaagtggcgcgaagcggttgcacacc	960
Db	43781	TGGCGGACAGGGTGGCGGGCGACACCCACTCAACAGGTGGCGGACGGGTGGCACACC	43840
Qy	961	cactgcaacaagcggtggcgcgaaggtgvcgtlaacacgcgcaatactactcgcgaattggccaa	1020
Db	43841	CACCTCAACAGCGGGTGGGAGGGGTGGGTATACCGCAATCACTCCGACATTGGCCAA	43900
Qy	1021	ccctaacgctacctcaagttacttgcgttcggtgtctgcgaacgcgaagttctaccgaagaac	1080
Db	43901	CCCTAACCGTACTCTAGGACTGAGTGGCTGGGTGCGCACCGCAGGTTCATACGAGCAAC	43960
Qy	1081	cgggcaaatcaaatgtgtgtgaagaacacccataaagttcggcgcgttgcgcgaagtttgaaag	1140
Db	43961	CGGCAAGATCAATGTGTGTAAAGACACCATCAAGGTGGCGGTGGCGAAGTCTTTGACGG	44020
Qy	1141	ccacggcgcaaccttcaactgctgcgcaacaatctatgggtlaacggaagacagggcgcaaatca	1200
Db	44021	CCACGGCGCAACCTTCACTGCGCGACAAATCTATGGGTATACGGAGACACAGGGGAAATCA	44080
Qy	1201	gaagcccaatgtctgaagcttgcgttgaagcgctacggttgaagaattgtaaccttgggtgaa	1260
Db	44081	GAAGCCCATGTGTCGAGTGGCTGTAAGAGCCCTGCTGTAAGATGTGTCAACTGGGTGAGAA	44140
Qy	1261	cgaaagtcgaatggtcatccacggtgaagaagccaaaacgctcaaggaatgaccattggaacgt	1320
Db	44141	CGAAGTTCGATGGCATATCCACGTATAAAGCCAAAAACGCTCGAGGAATCCATTGACAACT	44200
Qy	1321	gcatgcccaagaagcgttcggtgaagaacctgtaatacgttcaaaagcgagggagggcgcaagcgt	1380
Db	44201	GCATGCCCAAGAACGTTCGGTGAAGACCTGATTAACGGTCAAAAGCGAGAGCGCGCACGGGT	44260
Qy	1381	caataatctgaacatcaagaacagacagttgccaaggttcagacgacgaagaagttgtccagct	1440
Db	44261	CACATATCTGAACATCAACAGAACAGCAGTGCCTCAAAAGGTGCAGACGACCAAGTGTCCAGCT	44320
Qy	1441	caacgcacaacctctccttgaaatatgacacacttcaagggcgagcggttttggacaagttgt	1500
Db	44321	CACGCCCAACACTCTCTGTGAAATATGACAACTTCAAGGGCGACGATTTTGCCACGATGGT	44380
Qy	1501	tgcgcaccaacggttgcgaagcagtttgatgatactgagcatcgcgacttgaacgcatcgaagc	1560
Db	44381	TGCGACCAACGGTGGCGAAGCACAATTGTTGATACATAGACATCGAGTGTAAAGCGCATGCAAC	44440
Qy	1561	taaccacggaagttcgccttcggttgcgaagcgacagttgacatctgaagcttggcaaggg	1620
Db	44441	TAAACCAAGCAAGTTTGCCCTTGCTGTAAGAACCAAGTGCATGCAACTGGCAACGGG	44500
Qy	1621	caaatgcgcataagcgcgcggttaaaaacgcctacagataaaacccaggaatgaaaccaaa	1680
Db	44501	CAAACATCGCATACCGACAGTATAAACGCGCTATACGATTAACCCCAAGCATGCCACCA	44560
Qy	1681	caacgagcttgtaatccagacagaagtagcttgtaaaaaaaggggttggagctc 1729	
Db	44561	CACGAGGCTTGAATTCAGACAAAGTATGCTTGAATAAAAGGGGTGGAGCTC 44609	
RESULT	4		
AF037983	1833 bp	DNA	BCI 05-OCT-1998
LOCUS			
DEFINITION	Pseudomonas syringae hnp type III secreted protein (hnp) gene, complete cds.		
ACCESSION	AF037983		

```

VERSION      AF037983.1 GI:3695000
KEYWORDS
SOURCE       Pseudomonas syringae pv. syringae.
ORGANISM     Pseudomonas syringae pv. syringae
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
REFERENCE    1 (bases 1 to 1833)
AUTHORS      Charkowski,A.O., Alfano,J.R., Preston,G., Yuan,J., He,S.Y. and
              Collmer,A.
TITLE        The Pseudomonas syringae pv. tomato HrpW protein has domains
              similar to harpins and peptate lyases and can elicit the plant
              hypersensitive response and bind to peptate
JOURNAL      J. Bacteriol. 180 (19), 5211-5217 (1998)
MEDLINE      98422476
REFERENCE    2 (bases 1 to 1833)
AUTHORS      Charkowski,A.O. and Collmer,A.
TITLE        Direct Submission
JOURNAL      Submitted (10-DEC-1997) Plant Pathology, Cornell University, Plant
              Science Bldg. Rm 334, Ithaca, NY 14853, USA
FEATURES
source       1..1833
              /organism="Pseudomonas syringae pv. syringae"
              /strain="B728a"
              /db_xref="taxon:321"
              /note="pathovar syringae"
              115..1654
              /gene="hrpW"
              115..139
              /gene="hrpW"
              /note="hrp box"
              268..1593
              /gene="hrpW"
              /function="hypersensitive elicitor"
              /note="peptate binding"
              /note="HrpW, harpin"
              /codon_start=1
              /transl_table=11
              /product="hrp type III secreted protein"
              /protein_id="AAC62530.1"
              /db_xref="GI:3695001"
              /translation="MSIGINSTSYOPASTOLDPSALS GSKSPQNTFTSDESTAQVDP
              SALLFDIAROKDVSRFPDNTVQNPDTSSAATDPSNVVKKLSALVTSILQMLMLNK
              KODGODNEMODPRONGGGLTSPAESGDSGTQASGDSGGTTAATGGGGGTS
              PTTBEDGGSTPTAAGDGGGYSVSGAGSAGSAPSTEDGGGGSDGVTTPVTPOLAN
              PGRNSGNTVSDTTSLSPOSSEVNVKRTIYGAQVDPDGHATPTAAKSMGTGODE
              HOKPLELEAGVILKNVNGENADGITHVNAKNSQVITIDNVHAGNVSEDMITTAGEG
              GAKVTNLNITNSANGADKVIQJLADHILAVDGRKAVDFGLVFTNGSKQFPDDKSVL
              LINGVDATHGKFFALVKSDDSLKLAITGDIAMTDVKAHYDKTKASTOHTEL."
terminator   1615..1654
              /gene="hrpW"
BASE COUNT   448 a 555 c 529 g 301 t
ORIGIN
Query Match      42.9%; Score 741.2; DB 1; Length 1833;
Best Local Similarity 69.2%; Pred. No. 1.7e-133;
Matches 1111; Conservative 0; Mismatches 438; Indels 57; Gaps 5;
QY 150 tcttcctgctgcygtgacactgagtcgcagcgatagcatttcagttccttgcgtg 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 TTTTTCGTGGGGGTAGCGACTGAGTCGAGCGAAGGGGTTTGGTTCTTGGGGGG 60
QY 210 gttaggcataaaaaaagaactttaaaacagtgcgaatgagatgcccgcgaacagg 269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 AGGTGACTCTCCAGACGCGGTGCATCGGAGTA---TCGTCCGTCACAGACACAAACGGG 116
QY 270 aacgcgtgcctgctgcttgcacttgcagcgaagcaccacccaacacatccacatcc 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 117 AACCCATCCCGCCGTTGCTGCTACTCTTCGCGAACGCCAACGAGAGTCTCATACGAC 176
QY 330 ctatcgaa-cgagacgcgatacgcgcaactgtctgttaaaccttgaagctgctgct 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 177 CTAATGAACCCGACGAGGCGGGCGAGCAAGCGCGTAAATCCGGAAGACGACGCTGGG 236

```

```

QY 389 ccaattgccca-cttagcgaggtaacgcagcagatgacatcgccatca-----caaccgc 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 237 TCATATGTCATCCATCCATCGAGAGTAAGCACCATGAGCATCGTATCATCATCCACTC 296
QY 442 gccgcacagaccaccacccacccacatcgatatttttcgscgttaagcgcgaagttcctcaac 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297 CTACCAACCCCGCTTAACCCAGCTGAGCTTCTCGAGCGCTGAGCGGCAAGATGCCGACAC 356
QY 502 aaacacgttcgcgcgcgaagaacactcaagcgcgatacgcaccgcgattgattgttcgg 561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 CAATACGTTACACGATGAGAGACTGACAGAGGCGCTGACGCCGAGTCCCTGCTTGA 416
QY 562 cagcgacacacagaagaagcgttaacttcgycgcgcgcgcgcgcgcgcgcgcgcgcgcgc 621
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 417 CACTGCCACAGACGAAAGAGCTGAGCTTCGGACAGCCGCAATAGGTTCAAAACCCAC 476
QY 622 ggaagcgaagcgaagcccaagcagcagcagtcacacacacacacacacacacacacacac 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 477 GCACAGCAGCGCTGCACTGACACTGACCCCTCATCCACTGCTGCAAGCTGCTGACAGTCTGT 536
QY 682 catgctgtctgctgacatgctacccaactccaataaaagcagacaccacacacacacacac 741
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 537 CACATCGCTGCTGCAAAATGCTCATGACCTCAATAAAAACAGGACAGGCGTCAGGACTC 596
QY 742 gcttatagccagcgtccttccttcacagaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 801
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 597 CAACGAGTGGCAAGACCCGTTTCANAAACGAGGAGGACTCGGTACGCTACGCGCGAAGG 656
QY 802 cggggcgcgcggtatacgcgcgatacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 861
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 657 CAGTACGCTGCTGCTACGAGGAGCTTCAAGCGGTTACAGGCGGCGGACGAGGCGGACACGCGG 716
QY 862 aggcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 717 GACGGGTGCTGAGCGGTGCGGTGACATCCCGGACACCTGAAAGTGAGCGGCGGCTAC 776
QY 922 cagccacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 777 AACTCCAAACCGCTGAAAGCTGACGCGCGCGGTAGTATGCTGACATGCGCGCAGAGCGGAC 836
QY 953 -----ggcacacacacacacacacacacacacacacacacacacacacacacacacacac 996
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 837 CGGCGCGCGCTTGACACGAGGACGAAACCGCGGAGCGGTGACATGATGCTGACATGATGATG 896
QY 997 gcaaatcaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1056
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 897 TCAAGTACACCCCGCAACTGCGCAATCCCGCGCAACTCGGCAACGGGACGCGGCTTCGGA 956
QY 1057 cagcgaggtcttacgcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1116
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 957 TACACAGGCGCTCTCTTTCAGCAATCCGCGGCAAGTCAATGTCGTGAAAGACACCATCAAGT 1016
QY 1117 cgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1176
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1017 CGGTGCTGCGCAGGCTTTCAGCGCCACGCGTCAACTTTCACGCGCACAACTTCATCATGAGG 1076
QY 1177 taagagagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1077 CACCGCGCAGCAGGAGCAATCAGAAAGCCCTGTTGCAACTGCGGAGGCGGCTGTGCT 1136
QY 1237 gaagaatgtgaacccctgggtgagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1137 GAAAAAGCTGAACCTGGGGGAGACAGCAAGCCACAGCGCATCCAGTGAATGCCAAAAACCTC 1196
QY 1297 tcaggaagtcacacatgtaacacgltgcatgcccagaagctcgcgttgaagacgtgattacggt 1356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1197 GGAAGCAAGTCACATTTGATTAACGTGAGCGCCAGAGACGTCGCTGAAGACATGATACGCT 1256
QY 1357 caaagcgagggagggcgagcggtcaactaattgaacatcaagaacagagcggtgccaaggg 1416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1257 CAAGGCGGAAGCGCGGCAAGAGTCAACCAACTGAACATCACTAACAGCAGGCGCCAACG 1316

```


Oy	1417	tcgacgcgcgaaggttgcagcttaacgcgaacactctttaaataatgcgaacattcaa	1476
Oy	1417		
Db	1317	tgcgcattgcacaaagctatttcagctcattatgccgattaccocacttaaaagtccacgctttc	1376
Oy	1477	gaccgcagatttcgcgcagatggttcgcaccaaacygttgcgaacagtttgaatgacatgag	1536
Db	1377	ggccacacccgatttttcgtnacccttgcttcgcgcaccaattgggtgcacacagtttgatgacatgag	1436
Oy	1537	catcgagctggaacgcgatactgaagcttaccacgaacgagttgcgcctggttgaagaagcgacag	1596
Oy	1537		
Db	1437	tgttcgagctcattatggcgttgacccgcgcacccatggcgaatttgcactctctcaaaagccgacag	1496
Oy	1597	tgcagatctgaagcttgcgaacggygacaacttcgcacatgacacgaatgcgaacgctcaaacacgctc	1656
Db	1497	tgacgattctgaaacttgcccaactggcacaacatgccatgacacgagatgaacgacacgcctatg	1556
Oy	1657	taaaacccgacatgcagcccaaacacgcgaactttaaattccagaca	1702
Db	1557	caagaccaaagggctcttaccacgacacccggactctgtatgacagaca	1602

RESULT	5	
PSU16119		
LOCUS	710 bp	DNA
DEFINITION	Pseudomonas syringae pv. tomato BCT 29-AUG-1995 transcriptional units IV and V, promoter region, and orf, complete cds.	
ACCESSION	U16119	
VERSION	U16119.1	GI:563240
SOURCE	Pseudomonas syringae pv. tomato.	
KEYWORDS	Pseudomonas syringae pv. tomato	
ORGANISM		

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 710)	Loranc, J.M. and Keen, N.T.	Characterization of aVTE from <i>Pseudomonas syringae</i> pv. <i>tomato</i> : a hrp-linked avirulence locus consisting of at least two transcriptional units	Mol. Plant Microbe Interact. 8 (1), 49-57 (1995)

JOURNAL
MoL. Plant Microbe Interact. 8 (1), 49-57 (1995)
MEDLINE
95290716
REFERENCE
2 (bases 1 to 710)
AUTHORS
Keen, N.T.
TITLE
Direct Submission
JOURNAL
Submitted (18-OCT-1994) Noel T. Keen, Plant Pathology, University
of California at Riverside, Riverside, CA 92521, USA

```

source
1. .710
/organism="Pseudomonas syringae pv. tomato"
/db_xref="taxon:323"
299. .676
CDS

```

```

/ncode_size=100000
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="unknown"

```

```

/translation="MULTISKLNPKHPHYPTDSDATCSGKPMWSMRSNCPLEVTQ
HEHHHTTPAATDHHARFEFGAKROESSKTHVRRAEHSASDPECTVVRQHTTERQLR
HARQHPRESASGRQDAQRPVQHR"
BASE COUNT      181 a      198 c      185 g      146 t
ORIGIN

```

Query Match	40.0%	Score 692.2	DB 3	Length 710
Best Local Similarity	99.68%	Pred. NO. 6.2e-124		
Matches 694	Conservative	0	Mismatches 3	Indels 0
				Gaps 0

[illegible]

Qy 61 ctgagtcgcgaagatttcgttgaataaggtgtgtactgtgcatcttgtgcatattcaagg 120
|||||
Db 74 CTGAGTCGCAGATTTCGTTGATGAAGCGTGTGTACTGTCATCTGTGCAATTTCAAG 133

QY 121 cctcgcagtcgcgtgcgcgcacataccaagctcttcctgcgcgcgtgtgcacactgagtcgc 180
|||||
134 CCTCTGAGTCGCGGTGCGGACGACATACCAAGTCTTCTCTGCTGGCGGTGTGCACACTGAGTCGC 193
Db

QY 181 aggcctagcattccagttcccttcggtggcatataaaagaacclttaaa 240
| | | | |
Db 194 AGGCATTAGGCATTTCTACTTCCTTCCGTGGTGGGCAATAAAAAAGGAACCTTTTAAA 253

241 acagtcgcaatgagatccggcaaacgggaacccgctcgctggcgttgccactcaccttcg 313

314 AGCAAGCTCAACCCCAACATCCACATCCCTATTCGACGACAGCGATACGGCCACTTGC 373

Db 374 TCTGTAACCTGGAGCTGGCGTGGGICCAATTGCCACTTAGCGAGGTACGCACCAT 433

Db 434 GAGCATCGGATCACACCCCGCGGCAACAGACACACGCGCGTTCGATTTTCGGCGCT 4933

Db 494 AAGCGGCAAGACTCTTCACCAACACGTTGCGGAGCAGCAACACTCAGCAAGCGATCGA 553

Db 554 CCCGATGCACCTGTTGTTCCGCACGACACACAGAAGCGTCACTTCGCGACGCCGA 613

Db 614 CAGCACCCTCCAGAAATCCGCAAGACGCCAGCAAGCCCAACGACAGACGCCAGTCCAAATCGC 673

QY 661 taaatgatcatgctgcattgatcatgtcgttgcgtcag 697

Db 674 TAAATTGATCAGTGCATTGATCATGTCGTTGCAG 710

U94513	U94513	2708 bp	DNA	BCT	03-OCT-1998
LOCUS	U94513	2708 bp	DNA	BCT	03-OCT-1998
DEFINITION	<i>Erwinia amylovora</i> disease-specific operon, partial sequence; a				

ACCESSION U94513
VERSION U94513.1 GI:3414585
KEYWORDS type III protein secretion; Hrp pathway; hypersensitive response

SOURCE	Erwinia amylovora.
ORGANISM	Erwinia amylovora
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae	
Erwinia	

REFERENCE	AUTHORS	TITLE
1 (bases 912 to 2255)	Kim, J.F., Zumoff, C.H. and Beer, S.V.	HPwM, a new harpin of <i>Erwinia amylovora</i> , is a member of a family of plant cell wall-degrading enzymes called pectate lyases

ACCOUNT	REFERENCE	AUTHORS	TITLE
2 (bases 1 to 2708)	Kim, J.F. and Beer, S.V.		HPw of Erwinia amylovora, a new harpin that contains a domain

JOURNAL
MEDLINE
REFERENCE

J. Bacteriol. 180 (19), 5203-5210 (1998)
98422475
3 (bases 1 to 2708)

TITLE Direct Submission
JOURNAL Submitted (18-MAR-1997) Plant Pathology, Cornell University, 3

QY	1559	gctaaccacgcggaagttcgcgccttggtgtaaaaacgcacgacgtacgtacgtgaagtcgcaacg	1618
	11		
	1739	GCAGAGACGCGTAATTCTTCCTTCGTTAAACCGATTAGCGAGGGCTTAACGTCAATACC	1680
QY	1619	ggcaacatcgcacatgaccgcgcgtcaacaacgcgctacga	1656
	11		
Db	1679	AGTGATATCTCAGCTGGGTGATGTTGAAAAACCACTTCAAA	1642
RESULT	8		
AF083620/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			
CDS			
RBS			
CDS			
FEATURES			
Source			
gene			

Query Match	Best Local Similarity	8.9%; Score 153.6; DB 1; Length 4291;
Matches 329; Conservative	0; Mismatches 234; Indels 15; Gaps 2;	
1079 gccggcaagatcaatgtgtgtaaaagacacatcaagtcgctgcgtgcgaagctcttgcac 1138		

Db	3178	GGCGGTAAATCAGACGGGTGCTGTGATATACACCAATTACCGTGGAAAGCGAGTCAGTGTATTAT	3119
QY	1138	ggcagcgcgcaaccttaactgcccgaacaatctatggglaacgagaccagcgcaaat	1138
Db	3118	GGCAAAAGACAAACCTTACCGCCCGGTTTCAGATTATGGCGGATGGGGCCACTTCGAAAMC	3059
QY	1199	cagaagcccatgttcgagctgagctgaagcgctaaagtgtgaagaatgtgaactgtgtgag	1258
Db	3058	CAGAAACGGCTGTTTATATCTGGAGACGCTGCCACCTGAAAAACGTACCATGGGCCAC	2999
QY	1259	aacgagtcgatgcatgcaccagctgaagcaacaaacgcctcagaagtcaccattgacaac	1318
Db	2998	GACGGGGGGAGTAGTATTCATCTTTACG-----GTGATGCCAAATATGACNAT	2951
QY	1319	gtgcatgtcccagaagaagtgcgtgtgaagacctgattaaagtcaaagcgagcgagcgcgacgcg	1378
Db	2950	CTGCACGTTTACCAACGTTGGGTGAGGACCGCATTCACGTTGAAGCCAAACAGCGCGGGCANA	2891
QY	1379	gtcataactctgaacctaaagaacgacgcagctgccaaggttgacagacgacaaggttgccag	1438
Db	2890	AAATCCCAAGTTGAATACATACATACATGTTCCCTTTCAGACACGCGCTCGACAAATCTGTGAG	2831
QY	1439	ctcaacgccaacactactctgaaatctgaacaactcaagacctcaagcgcgacgattctgcgaagtg	1498
Db	2830	CTGAATGCGCGATACATACACCTGACCGCTGTACCAACGTAAGGCCAAAGACTTTGTATCTTTT	2771
QY	1499	gtctgcaccaacagctgycgaagcaagtttatgacatgagcatctcgagctftaaagcgatcgaa	1558
Db	2770	GTACCCACTTAACGGGGGCTCAACAG---GTTAACTGGGATCTGAAATCTGAGCCATATTCAGC	2714
QY	1559	gttaacacacgcaagttctgcacctgtgtgaagaacgacgtgaacatctgaactgtgcacag	1618
Db	2713	GCAGAAAGACGGTAAGTCTTCGTTCTGTTAAAGGCGATACGAGGGGCTTAAAGTCAATACC	2654
QY	1619	ggcaacatctgcgatgacgcgaagctcaacaacgcgctacga	1656
Db	2653	ACTGATATCTCACTAGTAGATGTTTAAAAAACCATCAACAA	2616

RESULT	9
DROMINI	
LOCUS	
DEFINITION	DROMINI 727 bp DNA INV 26-APR-1993
ACCESSION	Drosophila mauritiana minisatellite gene sequence.
VERSION	M62837
KEYWORDS	M62837.1 GI:157893
SOURCE	
ORGANISM	Drosophila mauritiana adult (strain E25H) DNA, clone lambda D4.2
REFERENCE	Drosophila mauritiana
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
TITLE	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
JOURNAL	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
FEATURES	1 (bases 1 to 727)
source	Jacobson, J.W., Guo, W. and Hughes, C.R.
	A Drosophila minisatellite contains multiple Chi sequences
	Insect Biochem. Mol. Biol. 22, 785-792 (1992)
	Location/Qualifiers
	1..727
	/organism="Drosophila mauritiana"
	/db_xref="taxon:7226"
	/dev_strage="adult"
	26..472
	/note="minisatellite 4.2"
	185..192
	/rpt_type=direct
	/note="Chi sequence"
	/number=1
	218..225
	/note="Chi sequence"
	/number=2
	237..264
	/note="Chi sequence"
	/number=3

	misc-feature	323.	.330	/note="Chi sequence"	
		/number=4			
	misc-feature	356.	.363	/note="Chi sequence"	
		/number=5			
	misc-feature	389.	.396	/note="Chi sequence"	
		/number=6			
BASE COUNT	124 a	143 c	276 g	184 t	
ORIGIN					
Query Match		3.9%;	Score 67.4;	DB 6;	Length 727:
Best Local Similarity		56.6%;	Pred. No. 0.0033;		
Matches 125; Conservative		0;	Mismatches 96;	Indels 0;	Gaps 0
Dy	773 ggcgagctcgtgacaccgcgcggccgaatagcggaggcgccgtgatcacaccgagtgcagact	832			
Dd	92 GGAGGAGGTGGTTCAACTTCACCTGGTGGCGGGAGGAGTGCGTTCAACTTCACCTGGT	151			
OY	833 ggcgagcgcggtgatcaagccaagcgcacaagcggtgtcggcggtgatatactccgaccgca	892			
Ddb	152 GCCGGGGGAGGAGGTGGTTAACTTTAGGTGGTGGTGGTGGTGGTGGTGAACAATGTA	211			
OY	893 acagcgagtgtagcgagcgcgttgcgcggcgacaccccattcaacaagaagtgcgcagcggt	952			
Dd	212 GGTGGAGCTGGTGCTGGTGGTGGCGGTGGTTCAACTTTGGTGCTGGTGGTGGCGGT	271			
OY	953 gccacaccactgccaaagcggtgcgaagttgcgcgtaac	993			
Dd	272 GGCCTAACAGTTGGTGCTGGTGGAGGAGGTGCCCTAAC	312			

RESULT	10
AB010070	AB010070 78172 bp DNA PLN 27-DEC-2000
LOCUS	
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 5, pl clone:MBK20.
ACCESSION	AB010070 BA000015
VERSION	AB010070.1 GI:2760166
KEYWORDS	.
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1 clone:MBK20.
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (sites)
REFERENCE	Sato,S., Kaneko,T., Kotani,H., Nakamura,Y., Asamizu,E., Miyajima,N and Tabata,S. Structural analysis of Arabidopsis thaliana chromosome 5. IV. Sequence features of the regions of 1,456,315 bp covered by nineteen physically assigned P1 and TAC clones DNA Res. 5 (1), 41-54 (1998) 98290546
TITLE	2 (bases 1 to 78172)
AUTHORS	Nakamura,Y. Direct Submission Submitted (07-JAN-1998) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 153-2, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yinakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934) Address for correspondence: kaos@kazusa.or.jp
JOURNAL	For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MBK20 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/grail-1.3/), GENSCAN (Chris Burge, MIT, http://CC-081.mt.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
COMMENT	

Denmark, <http://www.cbs.dtu.dk/services/NetGene2/> and
SplicePredictor (Volker Brendel, Stanford University,
<http://agrimini.zool.iastate.edu/cgi-bin/sp.cgi>).
Genes encoding tRNAs are predicted by tRNAScan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
<http://genome.wustl.edu/eddy/tRNAScan-SE/>).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is T21 and the 3' clone is MM12.

FEATURES

SOURCE

CDS

CDS

CDS

CDS

```
Location/Qualifiers
1..78172
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MBR20"
/clone_1b="Mitsui PI"
214..837
/note="emb|CA87947.1"
gene_id="MBR20.1"
/codon_start=1
/evidence=not_experimental
/product="transcription factor-like protein"
/protein_id="BAB11436.1"
/db_xref="GI:9759579"
/translation="MASPESSDLEAIOSHLEDLVCGDFMGDFPDPAFVGIMCI
EPHYKQEPDVPVLDPSFVNEFIOVGESSSSSPENSSSYETDOSKKAERPE
EEVARRHRCVRRRPMKFAEIRDPKAKGSRIMLGTFRESYDARADCAAFILGR
KAUINFLDACKYAPANSGRRKRSVDHELOTRQSSSSSCDAF"
complement(join(2200..2567,2679..2786,2884..2956,
3043..3219,3367..3855,4103..4219))
/note="gene_id:MBR20.2"
/codon_start=1
/evidence=not_experimental
/product="WD-repeat protein-like"
/protein_id="BAB11437.1"
/db_xref="GI:9759580"
/translation="MEETFAEYKOTGPCPCFSPNSRYAVANDRYLRITFSEVOYVOLT
SCIDISYIEMALDSEYITGLYKPKPMQAMSLQEMWCTCKIDGPAGISVARRSPS
RHITTSFQRLRVMSLNTACVHVQPRKGSQVSPNODKRAITCTRDCKDYV
LISCQSWIKGSPFVDTLDLADLEMSPDSSIIVWDSPLETKAIECGIGVTSWSPC
GQFLAIGSYDMLRVNLHWKTAEPHLSTVRAPCSAAIFKEIDEPLODMSEL
DENEMPSYDASEGYISVREMEIPVAFPOKPADPNPKOGVGLAMGKDSQYIC
TRNSMPTALIMDMRLLEVAALIVOKEPRAAVWDPCTRLCTGSSHLVMTWPSG
AFQCNPLPGFSRIDKWNIDGSCILLKDKDAFCCATYPSLPBSSDYSDD"
complement(join(5140..5587,5725..5969))
/note="gene_id:MBR20.3"
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11438.1"
/db_xref="GI:9759581"
/translation="MFSFLIRLLEVKYVAVASIVFVPSGLTLAGTAVGLVTTP
LFIETPLVPAITAILTLTGFTTGALGATAALIRGKGVKSKNIPRIGAPTM
FAQSLTPKINYEGETFGSGWCKSPQATPMFSGYGTWTAMGGSPFGKGDGSGG
STAGSPEPAAGACAGAAAAGAAAAGAAAVPGIGASAGAGPAGAGTAPAP
TGKGGKAGSKK"
6556..7818
/note="emb|CA85517.1"
gene_id="MBR20.4"
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11439.1"
/db_xref="GI:9759582"
/translation="MSSCSRTRTKAPRSARRNGGSESSASATIVADIDVLIQILST
LPIKTLRFKRVSKRWLITLNPFSNKVIRKSNHPLPSGFLNSPREIKYKFSYLD
DATOIRSSSLPMTDHOIDMIIMOSNGLLCLCKSCASNNHFNNTNYYVNPYKQY
TLHOIGHILSLAFDPSRSPKVCFLRGRSNSSASDSEIYHIEVYNSMGGL
RRVDPVPSSTPIEFESYFWMGNKAVMMYGGSSRDCSDINIOEIKLIPDHEHD
EPIDPRILMFLDSQGLYIEVNOSSSLURAYEMBSNSSSVKTVNDLEPIAA
FPEMIRFETDRRIYASVIGFVKEEDDAASYILHLIPQAVKYNFDKTKFKICDP
KSLVNDABEDHFYRFTFOFKISLVAN"
```

CDS

tRNA

CDS

CDS

CDS

```
complement(join(7889..8469,9056..9554))
/note="contains similarity to receptor-like protein
kinase
gene_id="MBR20.5"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11440.1"
/db_xref="GI:9759583"
/translation="MRSKLITLLIGVSLAFIVIPFIFIRKRESSSTESDOYVE
SLDHNRGFSSETEFVITFOGGEITLDCIDLDAGEYIGSSYGTILKASLORSGKR
VLRFLRVCYTRSDSKENGTIFELGVRHENVPLGLFAGNNGELAMHPFGSGN
LISDFIRSGDSESRWINILRTTIGISALDHLRGMOKPIVHGMUKSNVLSFEP
RISDFGLHLINLSAGOEIIDVSAEYKAPPELLKMDKVESDYSISGVTMELVSG
KEPVENATGDEFEYLPDFMKNAYVLDRLSLYRPEILGSDDNISECVALKYFQLAMS
CCSPSPSLRPVYKVLKLEBIGKF"
11636..11722
/note="gene_id:MBR20.6"
/evidence=not_experimental
/product="tRNA-Leu(CAA)"
join(12520..12539,12603..12763,12892..13043,13128..13246,
13356..13447,13592..13681,13826..13897,14025..14162,
14266..14421,14707..14845,14945..15011)
/note="gb|AAD38624.1"
gene_id="MBR20.7"
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11441.1"
/db_xref="GI:9759584"
/translation="MOPALIPDGPSEENVTRILKAVMTVPIGIAITTAICJFVLMW
QNLSSYPYAOAIWIGHFACVLELMABEPLYITLQTLRLIVETATLARCVTLY
SLVAMOTNMEKVIIFALSQVAYGSLFTGWAYVLLIGVYSSNLPFPRGNEFDEN
DLASKMKLTFPQSFRRKLILOEVLKTPGSIYRVWVFLPPESSYTIIPAFASGDYEN
KRLCIYLVAKVIYILIGLIFPMARGPSYSTILILYGERKSDSEALATOFYLYII
VLAMNGSEAFLEHVAVGTRKNELENSNDMLIFSLIVADNILLISACATIGITMANSIS
WOILLISITLISEKTLDRKNWEFFPLHFGIGVICFLSAIVIYRREKVFTRIKR
RFEDYNHDD"
join(15335..15610,15690..16364)
/note="contains similarity to RING zinc finger protein
gene_id="MBR20.8"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11442.1"
/db_xref="GI:9759585"
/translation="MVCYTFWLVCYKGLVSEVTRIDSKOIGFVYALICDHEDRLYE
MKVILGSESTHQAAELAAIILHNLNMALEFLKAVYFPQDDSNILRQKAPRNS
TVATLVKEVSLQSKFSFCEALPYMKDITFELKAKQALISQILNRGSDVYMETCPIC
YEHVTSDEKEFVPGCFHRFCFDICIKQADVALFEAKVAVNCPFGCSSELQREDCGV
LKPKQDRMTWYKRAKSMIKAKVLDIVFCCTCDNMAKVDLLEYTKTFEVAELSGVK
CTECGVCFCGCRAGMWSGMTCEYFPRSENPSPEDVEYR"
complement(join(16863..16984,17088..17154,17272..17340,
17462..17578,17642..17731,17790..17864,18056..18137,
18232..18282,18422..18574,18643..18684,18782..18880,
18966..19037,19148..19216,19525..19593,19716..19812,
19908..19979,20055..20122,20745..20826,20904..20954,
21053..21133,21251..21292,21434..21532,21630..21734,
21845..21917,22031..22302))
/note="gene_id:MBR20.9"
p1111717454
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11443.1"
/db_xref="GI:9759586"
/translation="MSLVEISGDMAAMPNGRVPPPPRPVPRRLPWFQAFDHT
GAGVWVGPFRPAKRRLSLKPLHWKTIISDQSLMDELQHRHGSQTAILEDISELT
LFPEVAREKIRLHDLRLRASVRYNVSQYWRANKYINISMPIDMTWVLADSEV
VDVOIKELIKFCPTNEMELKTYTDKAKAGREYQYLLELMKVPDEAKLRYFSK
TOFGTKITELKEKLVVTSACEVRSSEKLEIKKIPICLGNTNQGDRGVGLDSS
VDNHVYKSMHYCYKLVASSESLDVKYKDIQSLKSIQVSLAQNIOALIKRLK
LKQELTASEIDGPASVFCNTLKPFIISAEEMATYSLISVYLFKAHBNYKQAL
EKKRAMQOIDIRANDIEMLTKTVNITPLADMAVLDMDEVLDVDDIENILRCPK
EEMELLNVTGDKATLGKCDQLAKAKAPLKEHFRVINAFLPSLTPYVLEWKYGVES
```


	gene	LINVTQALHATPOVRGAVGVTSLSALIRPDEAGLGLAHAKKSWYSLSEAPAEVFN KLIVDDQNQLRAALDARRAALSISGIDVDPAQITSGVADNRRKEFGALSILNVLAN LNREDRDIITEALKRLPLPYATTLGVSVGSGFNNVSYVLDPPLVAIVDFVPQPKL PDSIADYLRGFIOERMTIRKSP" complement(5272..6324) /gene="RV3498c" complement(5272..6324)
CDS		/note="RV3498c", (MYO23_05c), len: 350. Unknown but similar to Mycobacterium tuberculosis proteins MTC128.10 (346 aa), MTC119H5_32c (346 aa) and MTVO51_05. Hydrophobic region at N-terminus. FASTA scores: 297050 MTC128.10 (346 aa) opt: 755 z-score: 883.6 E(): 0; 36.8% identity in 340 aa overlap; and 297182 MTC119H5_32 Mycobacterium tuberculosiscosml (325 aa) opt: 699z-score: 820.1 E(): 0; 47.0% identity in 249 aa overlap. Tbpase score is 0.878"
		/codon_start=1 /transl_table=11 /product="hypothetical protein RV3498c" /protein_id="CAAI7735.1" /db_xref="GI:2924435" /db_xref="SPTREMBL:O53543"
Query Match	3.7%, Score 64.4; DB 3; Length 47852;	
Best Local Similarity	47.5%; Pred. No. 0.0054;	
Matches 191; Conservative	0; Mismatches 211; Indels 0; Gaps 0;	
Oy	792	cggcgcatagatgagggggcgggttaacacgatgtgcagacgtgcccgcgggtatcacgc 851
Db	20649	CCGGCGGCACCGCGCGGCCCAAAGCGCGCGCGCATGGGTGGCTCCGGTCTGTATTATG 20708
Oy	852	caagcgacaacaggcgggttgccgctgaactaccgccccaacaggcgggtggccagcg 911
Db	20709	CCAGCGGGATTGGCCCGCACGGCGCGCGGTGGAGACTGCGGTAAACGCCGCGCGCGCGC 20768
Oy	912	gtggcgagcgcaacacccactgtgaacagtgtgcgcgagcggtggcaacaccactgcaacag 971
Db	20769	GGGCGCGGGGGCGCCCGGCACCGGAAGAACCAGGGGGGGGTGTGTGCGCGCCCGCGGCAAGG 20828
Oy	972	gcggtgtagaggtgtgcgttaacacccgcacaatactccgcagtttggccaacacctaacgtta 1031
Db	20829	CCGGTAGTGCGCGGCACCGCGGCGCAAAGCGGCGCGCGGCGGCGGCGAGCGCGGCAACGG 20888
Oy	1032	cctaaggtactggttcggtgtgtcgcgaacccgcaagtttaccgcagagacgcgcgaagaatca 1091
Db	20889	ATGCAACCGCTAACCGGTGCCACCGCGGCGCACCGGGTTTTCCGTGGAGACCGGCGGGCGG 20948
Oy	1092	atgtgtgtaaaagacacatacgaatgtgcgcgttgcggaagtctttgacgcccagcgqcaa 1151
Db	20949	GCGGGCGCGCGGCGGCACAACCGGGGTGGCGGCGCACCAACGCGTCCGCGCGCGGCAAGCGGCA 21008
Oy	1152	ccttcactgcgcgacaatatctatgggttaacgagagacaggcg 1193
Db	21009	CCGGCGCGCGCGGCGGCGCGCGGTGGTGTGCTGGTGTGTCGCGG 21050
RESULT 12		
LOCUS	SC2G38	40549 bp DNA BCT 13-DEC-2000
DEFINITION	Streptomyces coelicolor cosmid ZG38.	
ACCESSION	AL445503	
VERSION	AL445503.1 GI:10803131	
KEYWORDS	ABC transporter ATP-binding protein; ABC transporter transmembrane protein; asnc-family transcriptional regulatory protein; hydrolase; integral membrane protein; lipoprotein; marr-family transcriptional regulator; oxidoreductase; iron-sulfur binding protein; oxidoreductase, molybdopterin binding subunit; regulatory subunit; secreted deacetylase; secreted lyase; secreted protein; tetracycline transcriptional regulator; two component system histidine kinase; two component system response regulator; ung, uracil DNA glycosylase.	
SOURCE	Streptomyces coelicolor.	
ORGANISM	Streptomyces coelicolor	

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE
AUTHORS
1 (bases 1 to 40549)
Redenbach, M., Klier, H. M., Denapalte, D., Eichner, A., Cullum, J., Kinash, R. and Hopwood, D. A.

TITLE
A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome

JOURNAL
Mol. Microbiol. 21 (1), 77-96 (1996)

REFERENCE
AUTHORS
2 (bases 1 to 40549)
Saunders, D. C. and Harris, D.

REFERENCE
AUTHORS
3 (bases 1 to 40549)
Unpublished

REFERENCE
AUTHORS
Cerdeno, A. M., Parkhill, J., Barrell, B. G. and Rajandream, M. A.

TITLE
Direct Submission

JOURNAL
Submitted (13-Oct-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT
Notes:
Streptomyces coelicolor sequencing at the Sanger Centre is funded by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Framemap program of Bldp et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2638 .

FEATURES
source
1. 40549
Location/Qualifiers
/organism="Streptomyces coelicolor"
/db_xref="taxon:1902"
1. 40549
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 2638"
complement(1..320)
/gene="SCG38.01c"
complement(<1..320)
/gene="SCG38.01c"
/note="SCG38.01c, possible regulatory protein(fragment), len: >106 aa; similar to N-terminal region of TR:CA01635 (EMBL:AL391072) Streptomyces coelicolor putative regulatory protein SC9M4.08, 116 aa; fasta scores: opt: 218 z-score: 308.8 E(): 9.3e-10; 46.2% identity in 91 aa overlap"
/codon_start=1
/transl_table=1
/product="putative regulatory protein(fragment)"

misc-feature
/protein_id="CAC13060.1"
/db_xref="GI:10803132"
/translation="MTPATAGTRELPHPTAEIRLSCVLAHSDPRLRYRDLA
DSGALFSCSHEDLPVTKSTTYHFRVLRSGVIRQTYRGTAAMGLRRDDLDIVPGL
LDAL"

misc-feature
1..93
/note="nominal overlap with Streptomyces coelicolor cosmid SC264"
complement(326..329)
446..452
462..451
/gene="SCG38.02"
462..1541
/gene="SCG38.02"
/note="SCG38.02, probable oxidoreductase, len: 359 aa; similar to SW:NEMA_ECOLI (EMBL:D86311) Escherichia coli N-ethylmaleimide reductase (EC 1.-.-.-) Nema, 365 aa; fasta scores: opt: 509 z-score: 564.1 E(): 5.6e-24; 32.1% identity in 364 aa overlap. Contains Pfam match to entry PF00724 oxidored_FMN, NADH:flavin oxidoreductase / NADH oxidase family"
/codon_start=1
/transl_table=1
/product="putative oxidoreductase"
/protein_id="CAC13061.1"
/db_xref="GI:10803133"
/translation="MSALFEPRLDITIPNIRMPICQYSAPEGSAGYPCDMH
AHGARAAGTGLIVETAVGSPEDRISPDQGLANDQVEAFRITFELRGQTPVA
VOLAHAGKRASTAPGKAGAPVAGDAYGQPLAPALAFDEHHPVPTETVYQIDAV
GFEADARRALAGFEIIEHGAHYLLHEFSPSNORDAYGSGYANRFFALEAV
DAVRWPDGKDPFFRVATMDLERGGKPPDPTVFARDLHAGDLDIVSGVAVPR
VRTPGRCQVYFARVAKGSLPLPAAGCLITPEQAEKILANGADAVLGRRELLRN
PSWAQARALEVDARMPQYIGWM"

misc-feature
468..1505
/gene="SCG38.02"
/note="Pfam match to entry PF00724 oxidored_FMN, NADH:flavin oxidoreductase / NADH oxidase family, score 318.10, E-value 1e-91"
1739..2539
/gene="SCG38.03"
1739..2539
/gene="SCG38.03"
/note="SCG38.03, probable secreted lyase, len: 266 aa; similar to TR:004701 (EMBL:M94651) Fusarium solani pectate lyase A precursor (EC 4.2.2.2) Pella, 242 aa; fasta scores: opt: 605 z-score: 664.1 E(): 1.5e-29; 43.6% identity in 236 aa overlap. Contains possible N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=1
/product="putative secreted lyase"
/protein_id="CAC13062.1"
/db_xref="GI:10803134"
/translation="MTAVTRPARARAVTGALGALISVGLMTSGASSAOATWPTPN
GSEGVSTLSVSGTDYGMKRLXRGDLSGGDDQDPIELAGVALIKVITGAP
ADGVHCKSGCTILQNWMEDEGDAITFPGSSSNVTVSGGAKRADKRVQFNAGT
LNIISGFAVKNFTFVRSNGCSTQRTIRINLNGIIVNMKGIGAIINNYGDSATLRN
ITIVGDSKRTIYPCQKXIGYDNDGDEPDSDNGSDGTCKYSSSDITYK"

misc-feature
complement(2544..3350)
/gene="SCG38.04c"
complement(2544..3350)
/gene="SCG38.04c"
/note="SCG38.04c, conserved hypothetical protein, len: 268 aa; similar to TR:053782 (EMBL:AL021943) Mycobacterium tuberculosis hypothetical 26.5 kDa protein MTV040.02, 240 aa; fasta scores: opt: 735 z-score: 848.5 E(): 0; 51.4% identity in 218 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAC13063.1"
/db_xref="GI:10803135"
/translation="MGARTVRAHDERGCAARMGSGDPGLTLRPLRSARVYLSLL

gene
CDS
complement(<1..320)
/gene="SCG38.01c"
/note="SCG38.01c, possible regulatory protein(fragment), len: >106 aa; similar to N-terminal region of TR:CA01635 (EMBL:AL391072) Streptomyces coelicolor putative regulatory protein SC9M4.08, 116 aa; fasta scores: opt: 218 z-score: 308.8 E(): 9.3e-10; 46.2% identity in 91 aa overlap"
/codon_start=1
/transl_table=1
/product="putative regulatory protein(fragment)"


```
/translation="MKRRAGTWAALLLAAYMLAGGNRPADQDSKAKQTEVEPVYITD
DASNDVITKEPKIVSLMPSTNETITLALGDKVVGVTINDTIPKVKAEKVDMD
NNVNAEKVIAKDLVLAHSSSASADAIKQDKAGITVLTVNDAQSEFEVKSIEMI
GGAGCAEKADQDLVKSMDLQIEKATQISKDEKSVFIEVSPDPDIYTKGJEFM
NMLNVIHAKNAADQGTWOMTDEAIVKLNDAIYTTGTAKAKAVERKDGMSSEINAV
KHHRVYDNDPLVTRSGPRLISGVEELASITPDIPFKE"
4343..5165
/gene="yyrd"
4343..5134
/gene="yyrd"
/function="unknown"
/translation="MDIOLKEKLVITSTSGIGRAAKSFLOEAGVAVTNGRKQETV
DRTIELSGYGVHGAADLSTDEAAATIEKYNEDGIDILYNNIGFEVDRADVT
DEEMQYFEVNWASAVRTSRHLPKMLANSRILNITASEACVKKPLPIWIPYSMTFA
LISLRGAEMTGTNTVNVSVLPQPTWEGVAAVMEGAQAQADPTDFIKDYERKN
EPTSLIORAYATGEAVANTIVFLASDAASAINCTAORVEGGIIRSL"
5145..5165
/gene="yyrd"
complement(5172..6050)
/gene="yyrE"
complement(5172..6050)
/gene="yyrE"
/function="unknown"
/translation="similar to RNA polymerase"
/codon_start=1
/transl_table=11
/protein_id="CAB15325.1"
/db_xref="GI:2635833"
/translation="MDVLEADPTRAVIGSGPLMDENGRFLWYDILSGSELIHFDPEK
INRSIKFSPVATALAKYSKDELIMTKDGFYLYHLPDLSLEKIKOPKDMHESIRFDEK
KCDPYGRLMAGITTSMEGEOKASLYRLINDGSLVKIKDOYSTNSGMDRERKLMTYI
DPTQELIVRYSTDPOSQDVSNPEPYRFPQSGLPDGMITDONGMLWALVFGSRYVA
IDPQKREINISISVPAKYVTCAGFGGRDLKTLIYTTATBOMTEKERYEOPHAGLFLSA
OLETGYPVPFAGDY"
complement(6122..7843)
/gene="yyrG"
complement(6122..7843)
/gene="yyrG"
/function="unknown"
/translation="similar to two-component sensor histidine kinase
[yyrH]"
/codon_start=1
/transl_table=11
/protein_id="CAB15326.1"
/db_xref="GI:2635834"
/translation="MPHFGOMLVITLITVNLVASFYIDARFSDAESNSGTSKATP
DLEAVLDYNEGTGEVNDPLKKSVDKQGMQIIDSBNTPYSSVPPDYQYTKK
ELLSTYIKTKLHNKILNTAINIEDSYILLGWSKSSQILFVEKRRQDLSLAHY
KSSITDIYIRKKGAITYLLDSNGKILDSINSTSEKRTAQMLELAKSPWYKRAIS
VKILNDRMVAIVNPVYVTDQEFNKSFLKYVKAMFLVMAVLFMYIIMYVWYFR
FGLPIFHTIRMLVNLSSKLEPRNREGSPVSKNKGKIKOPRFEGFIESDOLTE
TIRBDRNRKEKIOATPREEWIAGLSHDKLPLSLISYGMLESKOYDMPERSEGGDI
VYRSEKSEYKSLIEDNLNTRYLKNDAIPIERKILTSIIPFKVYIEPKRNPSEGGDI
SVRSKEHEIFALDEAFRRILLENLGNVAKHNGKTEIOVILIEQTKNHSIAVKQNG
KGMDEETITHLNRRYRGNTKDSITAGTGLGIAIAKELVHLNHTIHVNSRNIQTVI
TILFRKO"
complement(7861..8970)
/gene="yyrH"
complement(7861..8970)
/gene="yyrH"
/function="unknown"
/translation="similar to two-component response regulator [yviG]"
/codon_start=1
/transl_table=11
/protein_id="CAB15327.1"
```

```
/db_xref="GI:2635835"
/db_xref="SPTREMBL:P94504"
/translation="MDQGEOKRKQODETYDIEHLIACFSPIRIKKLSTSTQEREDL
EDELIKMEKADMLICDQVDPFWEFILITVWDENSIISHMKHKGALVCSFFAVLI
Query Match 3.3% Score 57.6; DB 2; Length 209510;
Best Local Similarity 50.5%; Pred. No. 0.082; Indels 27; Gaps 2;
Matches 215; Conservative 0; Mismatches 184;
QY 1090 caatgtgtgtaagacacacacaaagtcgcgcgtgcygaagctcttgagcgccacgcgc 1149
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190199 CAAGTGTGTGCGACGAGAAACATTTTCGTACCAAAAATFACAACTATGACGGGAAAGGACA 190258
QY 1150 aacctcaactgcgcacaactctatgggtgaacgagacgagcgaggaataagaagcccat 1209
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190259 GCGGTTTGTGTGCGAGGGAAGAAATTTAGGTGACGGAAGCCACTGCAAAAACCAACCTGT 190318
QY 1210 gtccgagctgtgcgaagcgctacgttgaagaatgtgaaccttggtgtagaagcaggtcga 1269
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190319 TTTTCGTTGTGAGAGATGAGACCAACCTGAAAATGTGTCTTGTTGTCACCTGCACTGA 190378
QY 1270 tggcatccacgttgaagccaaaacgcctcaggaagtcaccattgacaacgttgatgccca 1329
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190379 TGGCGTGCACACCTTATGGAAC-----GTTAACATTGCAAAATGTGAATGCGA 190426
QY 1330 gaacgtcgtgtgaagacctgattcagtcgaagcgagagcgagcgagcgatcaactatc 1389
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190427 AGATGTTGTGTGAGGATGCGTTTAAACGCTGAAGGAAGGAGGA-----AAAGT 190471
QY 1390 gaacatcaagaacagcagctgcacaaaggtgcagacgacgaacgtgttcagctcaacgcaa 1449
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190472 GAGCATGCAGCGCGGTTCTCTCAAAAAGCGTCAGATTAATATTCAAATCAATTAAGC 190531
QY 1450 cactcacttgaacacgcacaactcgaagcgacgaatgtgcagcagatggtgcagacaa 1509
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190532 CAGTACTTGCACAGTGAAATTTACAGCGGGAATATAGCGGGAAGTTCATTGACAGCT 190591
QY 1510 cgggtgg 1515
|||||
Db 190592 TGGTGG 190597
RESULT 15
E32986 10732 bp DNA PAT 07-FEB-2001
LOCUS E32986
DEFINITION Gene encoding cellulose synthesizer.
ACCESSION E32986
VERSION E32986.1 GI:13022340
KEYWORDS JP 2000060568-A/1.
SOURCE Vigna angularis.
ORGANISM Vigna angularis
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Vigna.
REFERENCE 1 (bases 1 to 10732)
AUTHORS Koichi,M.T.K.K. and Sato,D.S.
TITLE Gene encoding cellulose synthesizer
JOURNAL Patent: JP 2000060568-A 1 29-FEB-2000;
COMMENT KOICHI MIZUNO,MITSUJI GIYOUSAI SHOKUBUTSU BIO KENKYUSHO
OS Vigna angularis
PN JP 2000060568-A/1
PD 29-FEB-2000
PF 26-AUG-1998 JP 1998239998
PR KOICHI MIZUNO,TOMOHIKO KATO,SHIGERU SATO,DAISUKE SHIBATA PC
C12N15/00,A01H5/00,C07K16/40,C12N5/10,C12N9/10,C12P21/02// PC
(C12N15/09,C12R1:91),(C12N5/10,C12R1:91),(C12P21/02,C12R1:91), PC
C12N15/00,
PC C12N5/00,(C12N15/00,C12R1:91),(C12N5/00,C12R1:91) CC
FH Key Location/Qualifiers
FT CDS location/Qualifiers
(1)..(3375).
1..10732
```

/organism="Vigna angularis"
/db_xref="taxon:3914"
BASE COUNT 3149 a 1212 c 2074 g 2046 t 2251 others
ORIGIN

Query Match 3.3%; Score 57; DB 10; Length 10732;
Best Local Similarity 13.4%; Pred. No. 0.2;
Matches 155; Conservative 451; Mismatches 525; Indels 22; Gaps 3;

QY 371 cctggagcgtgctgctgctcaattgcccactagcgaggtacgcagcgtgcatcgcc 430
DB 9446 YGSGRASNARNGTSASNGNHSARAGVAMTTHRKGTTHRASRSHSGSNGCASHSYSGY 9505
QY 431 atcacaccccgccgcaaacacacccacacacacacacacacacacacacacacac 490
DB 9506 HSRFVAGHSHSRASNVAAASHSHSRHAAASRSRASAAAGYGRASASYSYTHSRCSAAG 9565
QY 491 agtccataac 550
DB 9566 SRTTRASRAASRASHSHARSGRGSYSYSGRCSYSYSSRAGYAAASRASASNMTRAS 9625
QY 551 cgtgtgttcggcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 610
DB 9626 HVAGMTGTSAAVAVASRVAGVSAATHRGNSRSRA-----ASRYSAAVATHRGASN 9676
QY 611 cagaatccgcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 670
DB 9677 SRSRGTHNGYTRHYSGRSRRHRGGGYSRVARASHSNNGNTHRTHRGHTRHRCVAA 9736
QY 671 agtcattgatcatgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 730
DB 9737 ASMTGRMTSRASHSRASRSRYSASNTHRCSRASHSYSGYTHRVARGYTRGNASVA 9796
QY 731 aatcgagcaacacccgctgctgctgctgctgctgctgctgctgctgctgctgctgct 790
DB 9797 VAYSTHRGMAHNSHVAHTRHSYSGTRASASHASYSVAHAARGHSAASNSHSHASH 9856
QY 791 tccgacgatacggcgcgcgctgacacccgatacgcagagtgccgcgcgctgatacg 850
DB 9857 ASSRBRAGGYRVAHSAHSTHGNASNRNGCYSHSCYSSRCSYSSRASRSRVAAASNR 9916
QY 851 ccaagcgaacacggcgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 910
DB 9917 SRRAAGYASNAASRSRYSARCGRSRNGASRSRYSGYAGAGRAAVALASTYRAS 9976
QY 911 ggtgcccggcgacacccacacacacacacacacacacacacacacacacacacac 970
DB 9977 RSRAASRNTSGTYARAGSHSTRTYASNSNGYMT-----KTHRGTYTRMTVAARGVAHGN 10031
QY 971 ggcggtgagcagagtgctgctgtaacacccgacacacacacacacacacacacac 1030
DB 10032 TRYSTRHSRGSRYVAGGHHSVACXYSTRASRSGYSTRHASTYRGVYASNHAAGYTHRH 10091
QY 1031 acctaggtactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1090
DB 10092 RAAASTRMTASNSHSCYSHSRGNASVSAASRSRMTARGASAAYSYGNHASTRASGTHRARCS 10151
QY 1091 aatggtgtaagaagacacatcagcgtcgcgctgcaagcttgcagcgccacggcgca 1150
DB 10152 RASGTGGASNGTHRSVAGYSGARHSHRAKGSASRSRR-----GNVATHRTHRSR 10203
QY 1151 accttaactgcccacaatctatggttaacgagacacagggcgcaaatcagaagcccat 1210
DB 10204 CYSGYHASASNGNSNGYGYRTYRASNGYSGHTRASYSASYSRRAAGSNGNSGVAGANG 10263
QY 1211 ttgcagctgctgtaagcgctacgttgaagaatgtaacctggtgtaagaagagtgat 1270
DB 10264 YSGNRAATHRASARGSGRMTASNGNSNGYSTRHASSRARGGNSRHYSGSRASNGY 10323
QY 1271 ggcatcaacgftgaagaagcacaacacgctcaaggaagtcacacatgacacgtgcat 1330
DB 10324 GNGASNGYSASNGNSASNSRASNSAGNHSHTHRCAGAGYSAAAHASASNTSYVASHRGGA 10383

QY 1331 aacgtcgtgtaagaacctgaattacggtcaaaaggcagggcgccagcgtcactaatctg 1390
DB 10384 SNYSASNSRASNSCYGGGGAATYCYGGGNGNSRMTSRYSBGCCSSRYSRSHASYSASYSR 10443
QY 1391 aacatcaagaacagcagtgccaaagtgcaagacaaagtgctcaagctcaagcccaac 1450
DB 10444 ARGASNASTRGTHRAAASRGYSAAAGCYSGNGCTHRASNGYSGNSAAMTAAARYSASA 10503
QY 1451 actcactgaaatcgacaactcaagccgcagcatttcggcagcagtgctgcaacccaa 1510
DB 10504 ASRHASRVAVAATHRGNHASNTHRVATHRASNASNTHRAATHTHRSTRHARGAAS 10563
QY 1511 ggtggcaagcag 1523
DB 10564 NVAASRRHSRY 10576

Search completed: October 4, 2001, 22:26:29
Job time: 7060 sec

Fri Oct 5 10:04:10 2001

us-09-597-513-1.rge

Page 23

... PAGE BLANK (USPTO) ...

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2001, 20:47:29 ; Search time 137.85 Seconds
(without alignments)
7875.531 Million cell updates/sec

Title: US-09-597-513-1

Perfect score: 1729

Sequence: 1 tccactgcgtgatttgaa.....tgaataaagggtgacac 1729

Scoring table: IDENTITY-MDC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 0601:*

- 1: /SIDSL/gcgdata/geneseq/NA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/NA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/NA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/NA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/NA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/NA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/NA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/NA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/NA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/NA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/NA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/NA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/NA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/NA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/NA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/NA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/NA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/NA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/NA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/NA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/NA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1729	100.0	1729	20	Hyper-sensitive res
2	1729	100.0	1729	21	Pseudomonas syring
3	1729	100.0	1729	21	DNA encoding a hyp
4	155.2	9.0	1344	20	Hyper-sensitive res
5	155.2	9.0	1344	21	Extinct amylovora
6	155.2	9.0	1344	21	DNA encoding a hyp
7	58.2	3.4	591	19	Bacillus sp. pecti
8	57	3.3	10732	21	Gene encoding a su
9	56.6	3.3	390	13	Randomising oligon
10	56.6	3.3	390	14	PCR primer for 5'
11	56.6	3.3	390	22	Sequence containin

12	56.6	3.3	591	19	AAV69879
13	55.8	3.2	666	20	AAV89484
14	54.2	3.1	538	19	AAV64546
15	54.2	3.1	538	19	AAV44437
16	54.2	3.1	538	20	AAZ19347
17	54.2	3.1	538	20	AAZ19347
18	52	3.0	114955	20	AAV53491
19	51	2.9	743	21	AAV35115
20	51	2.9	985	19	AAV64548
21	51	2.9	985	19	AAV44439
22	51	2.9	985	20	AAZ19349
23	51	2.9	985	20	AAZ19137
24	50.2	2.9	3183	20	AAZ26611
25	49.8	2.9	6530	14	AAO51557
26	49.8	2.9	6530	20	AAZ22072
27	49.8	2.9	6530	22	AAV68948
28	49.4	2.9	3946	18	AAV93610
29	47.8	2.8	7720	21	AAV53800
30	47	2.7	1028	13	AAO27091
31	46.8	2.7	732	21	AAV12333
32	46.8	2.7	3086	11	AAO06470
33	46.8	2.7	8045	21	AAV65171
34	46.6	2.7	407	19	AAV64539
35	46.6	2.7	407	19	AAV44430
36	46.6	2.7	407	20	AAZ19340
37	46.6	2.7	407	20	AAZ19128
38	46.6	2.7	10266	17	AAV33007
39	46.6	2.7	14704	13	AAO20685
40	46.2	2.7	717	21	AAV39930
41	45.8	2.6	888	17	AAV59269
42	45.8	2.6	1548	19	AAV33277
43	45.8	2.6	1581	19	AAV33278
44	45.8	2.6	2888	17	AAV59268
45	45.8	2.6	3015	20	AAZ32027

ALIGNMENTS

RESULT 1					
AAZ24347	ID	AAZ24347	strand:	DNA;	1729 BP.
AC	AAZ24347;				
DT	07-JUN-1999	(first entry)			
DE	Hyper-sensitive response elicitor hrpW dspe gene.				
KW	Hyper-sensitive response elicitor; dspe gene; hrpW;				
KW	transgenic plant; disease resistance; insect resistance; ss.				
XX					
OS	Pseudomonas syringae pv. tomato.				
FH	Key	Location/Qualifiers			
FT	CDS	419..1693			
FT		/*tag= a			
PN	WO9907207-A1.				
PD	18-FEB-1999.				
XX					
PF	24-JUL-1998;	98WO-US15501.			
XX					
PR	06-AUG-1997;	97US-0055107.			
XX					
PA	(CORR) CORNELL RES FOUND INC.				
XX					
PI	Alfano JR, Charkowski A, Collmer A;				
XX					
DR	WPI: 1999-167125/14.				
DR	P-PSDB; AAV97851.				
XX					

Bacillus sp strain
Bacillus lichenifo
M. tuberculosis im
Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
Human adenosine A1
Arabidopsis thalia
M. tuberculosis im
Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
Trypanosoma cruzi
Loricin gene. Ho
Nucleotide sequenc
Mouse loricin gen
Mycobacterium tube
Genomic DNA encodi
XRY26 probe. Homo
Aspergillus oryzae
Sequence encoding
Neurospora crassa
M. tuberculosis im
Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
Mouse SRY-related
PKS 741 insert con
Arabidopsis thalia
Streptomyces prist
Nucleotide sequenc
Nucleotide sequenc
Streptomyces prist
Human METH1 relate

PR New hypersensitive response eliciting (dspr) gene and protein
 PR useful for providing transgenic plants and seeds with enhanced
 PR growth, and insect and disease resistance

Claim 1; Page 37-38; 56pp; English.

CC This the DNA sequence of the dspr gene of *Pseudomonas syringae*
 CC pv. tomato DC3000 that codes for a 42.9 kDa hypersensitive response
 CC elicitor (HRE) polypeptide (see AM97851), termed HrpW. To identify
 CC any HRE-like genes in the *P. syringae* pv. tomato DS3000 DNA flanking
 CC hrpR, cosmid pCPR2357, which contains this region in vector pCPR47,
 CC was isolated. A series of subclones in pML123 were constructed and
 CC screened for 2 potential HR phenotypes: (1) the ability to promote
 CC tobacco HRE activity in *Pseudomonas fluorescens* cells carrying
 CC pCPR2274, a delhrpZ PHR11 derivative, and (ii) interference with
 CC the HRE activity of *P. fluorescens* cells carrying wild-type PHR11.
 CC No subclones had the first phenotype, but one, pCPR2373, had the
 CC second. Transcriptional unit V of this subclone contained the
 CC 1275 bp open reading frame encoding HrpW. The HrpW protein or
 CC isolated DNA molecule can be used to impart disease resistance to
 CC plants, to enhance plant growth and/or to control insects on
 CC plants. This is achieved by applying the HRE protein in a
 CC non-infectious form to plants or plant seeds. Alternatively,
 CC transgenic plants or plant seeds transformed with DNA encoding the
 CC HRE can be provided.

Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 other:

Query Match 100.0%; Score 1729; DB 20; Length 1729;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tccacttgcgtgatttgaatgacattatagaacgttcaggtgtgtaaatcag 60
 DB |||||||
 QY 1 tccacttgcgtgatttgaatgacattatagaacgttcaggtgtgtaaatcag 60
 DB |||||||
 QY 61 ctgagtcgcagattctgttaaggtgtgtactgttcatttggcttaacag 120
 DB |||||||
 QY 61 ctgagtcgcagattctgttaaggtgtgtactgttcatttggcttaacag 120
 DB |||||||
 QY 121 cctctgagtcggtgcgagcaataccagcttctctgtgcgtgtgcaactgtacg 180
 DB |||||||
 QY 121 cctctgagtcggtgcgagcaataccagcttctctgtgcgtgtgcaactgtacg 180
 DB |||||||
 QY 181 aggcataagcatttcaagcttctgtgtgtgtgtggtgcataataaagaactttaaa 240
 DB |||||||
 QY 181 aggcataagcatttcaagcttctgtgtgtgtgtggtgcataataaagaactttaaa 240
 DB |||||||
 QY 241 acagtcgaatgagatgcgcgcaaaacgggaacggtcgtcgtcttggcaacttcg 300
 DB |||||||
 QY 241 acagtcgaatgagatgcgcgcaaaacgggaacggtcgtcgtcttggcaacttcg 300
 DB |||||||
 QY 301 agcaagctcaaccaccaatccatccctatcgaaacgagcagatagggcaacttgc 360
 DB |||||||
 QY 301 agcaagctcaaccaccaatccatccctatcgaaacgagcagatagggcaacttgc 360
 DB |||||||
 QY 361 tctgttaaaccttgtagctgtgcgttcgaattgccaacttaggaggttaacgagat 420
 DB |||||||
 QY 361 tctgttaaaccttgtagctgtgcgttcgaattgccaacttaggaggttaacgagat 420
 DB |||||||
 QY 421 gagcttcgcatcacaccgcgcgcaacagacccacgcacactcgtatttctgcgct 480
 DB |||||||
 QY 421 gagcttcgcatcacaccgcgcgcaacagacccacgcacactcgtatttctgcgct 480
 DB |||||||
 QY 481 aagcggcaagagctctcaacaacacgcttcggcgagcaagaactcagcaagcgatcga 540
 DB |||||||
 QY 481 aagcggcaagagctctcaacaacacgcttcggcgagcaagaactcagcaagcgatcga 540
 DB |||||||
 QY 541 ccgagtgactgttgcgtgcgagcagacacagaaagacgttcaacttcggagcgcga 600
 DB |||||||
 QY 541 ccgagtgactgttgcgtgcgagcagacacagaaagacgttcaacttcggagcgcga 600
 DB |||||||

QY 601 cagcacgtccagaatccgcagacgcaagcccaacgacgacccatccatccgc 660
 DB |||||||
 QY 601 cagcacgtccagaatccgcagacgcaagcccaacgacgacccatccatccgc 660
 DB |||||||
 QY 661 taattgtcagtgatgatcatgtctgtctgacagatgtctacaaactccaataaaa 720
 DB |||||||
 QY 661 taattgtcagtgatgatcatgtctgtctgacagatgtctacaaactccaataaaa 720
 DB |||||||
 QY 721 gcaagacacacatcaggaacagcctcgatagccagctctctccagaanaagcgagct 780
 DB |||||||
 QY 721 gcaagacacacatcaggaacagcctcgatagccagctctctccagaanaagcgagct 780
 DB |||||||
 QY 781 cgttacacgttcgcccagatagcggtggcggtacacccgagtgagacaggtggtggcg 840
 DB |||||||
 QY 781 cgttacacgttcgcccagatagcggtggcggtacacccgagtgagacaggtggtggcg 840
 DB |||||||
 QY 841 cgtgtatccgcaaacgcaacagcggtggtggtggtatatactccgacccgcaacggtg 900
 DB |||||||
 QY 841 cgtgtatccgcaaacgcaacagcggtggtggtggtatatactccgacccgcaacggtg 900
 DB |||||||
 QY 901 tggcgagcagcgttggcggtgcgacacccactgcaacaggtggtggtggtggtggtg 960
 DB |||||||
 QY 901 tggcgagcagcgttggcggtgcgacacccactgcaacaggtggtggtggtggtggtg 960
 DB |||||||
 QY 961 cactgcaacagcgttggcggtggtggtggtggtggtggtggtggtggtggtggtg 1020
 DB |||||||
 QY 961 cactgcaacagcgttggcggtggtggtggtggtggtggtggtggtggtggtggtg 1020
 DB |||||||
 QY 1021 ccctaacgttacctgaagtgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
 DB |||||||
 QY 1021 ccctaacgttacctgaagtgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
 DB |||||||
 QY 1081 cggcaagatcaatgt 1140
 DB |||||||
 QY 1081 cggcaagatcaatgt 1140
 DB |||||||
 QY 1141 ccacggtcgaactctcactgcgcgaacaaactatgtgtgtgtgtgtgtgtgtgtgtgt 1200
 DB |||||||
 QY 1141 ccacggtcgaactctcactgcgcgaacaaactatgtgtgtgtgtgtgtgtgtgtgtgt 1200
 DB |||||||
 QY 1201 gaagccatgttcgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260
 DB |||||||
 QY 1201 gaagccatgttcgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260
 DB |||||||
 QY 1261 ggaagccatgttcgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320
 DB |||||||
 QY 1261 ggaagccatgttcgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320
 DB |||||||
 QY 1321 gcatgcccagaacgttcgtgtgaagaccttgattcaggttcaaaagcgagggagcgagcgt 1380
 DB |||||||
 QY 1321 gcatgcccagaacgttcgtgtgaagaccttgattcaggttcaaaagcgagggagcgagcgt 1380
 DB |||||||
 QY 1381 caataatctgaacatcaagaagaacagcagtgccaaaggtgcagacgcaaggtttgtcaact 1440
 DB |||||||
 QY 1381 caataatctgaacatcaagaagaacagcagtgccaaaggtgcagacgcaaggtttgtcaact 1440
 DB |||||||
 QY 1381 caataatctgaacatcaagaagaacagcagtgccaaaggtgcagacgcaaggtttgtcaact 1440
 DB |||||||
 QY 1441 caagcccaactcacttgaaatcgacaacttcaaggccgacgatttcgagacagatgt 1500
 DB |||||||
 QY 1441 caagcccaactcacttgaaatcgacaacttcaaggccgacgatttcgagacagatgt 1500
 DB |||||||
 QY 1501 tgcgaccaaaggttgcgcaagcagttgtatgacatgagcatcgagtcgaaacgcatcgaagc 1560
 DB |||||||
 QY 1501 tgcgaccaaaggttgcgcaagcagttgtatgacatgagcatcgagtcgaaacgcatcgaagc 1560
 DB |||||||
 QY 1561 taacacggcgaagttcgcctgt 1620
 DB |||||||
 QY 1561 taacacggcgaagttcgcctgt 1620
 DB |||||||
 QY 1621 taacacggcgaagttcgcctgt 1680
 DB |||||||
 QY 1621 taacacggcgaagttcgcctgt 1680
 DB |||||||
 QY 1681 caccgagcttgaatccagacaagtagcttgaataaagggtgtgactc 1729

|||||
Db 1681 caccgagcttgatccagacaaagttagcttgtaaaaaagggtgagctc 1729
RESULT 2
AAD00673
ID AAD00673 standard; DNA; 1729 BP.
XX
XX AAD00673;
AC
XX
XX 08-SEP-2000 (first entry)
DT
XX
XX Pseudomonas syringae hypersensitive response elicitor encoding dspe gene.
DE
XX
XX Hypersensitive response elicitor; environmental stress resistance;
KW plant; dspe gene; ds.
XX
XX Pseudomonas syringae.
OS
XX
XX
FH Key Location/Qualifiers
FT CDS 419..1693
FT /tag= a
FT /product= "Hypersensitive response elicitor"
XX
XX WO200028055-A2.
XX
XX 18-MAY-2000.
XX
XX 04-NOV-1999; 99WO-US26039.
XX
XX 05-NOV-1998; 98US-0107243.
XX
XX (EDEN-) EDEN BIOSCIENCE CORP.
XX
XX Wei Z, Schading RL;
XX
XX MPI: 2000-376566/32.
XX
XX P-PSDB: AAY71098.
XX
XX
XX Application of a hypersensitive response elicitor protein to plants to
XX
XX impart stress resistance
XX
XX
XX Disclosure; Page 24-25; 84pp; English.
XX
XX
XX The patent discloses a method to impart stress resistance to plants by
XX
XX applying a hypersensitive response elicitor in a non-infectious form to
XX
XX a plant or seed. The present sequence is dspe gene encoding
XX
XX hypersensitive response elicitor protein from Pseudomonas syringae.
XX
XX The present sequence is used to transform transgenic plant or plant
XX
XX seeds to impart stress resistance.
XX
XX
XX Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 other;
SQ

Query Match 100.0%; Score 1729; DB 21; Length 1729;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 acagtgcaatgagatgcgagcaaaacgggaaacgggtcgctgctgttgccactacttcg 300
|||||
Db 241 acagtgcaatgagatgcgagcaaaacgggaaacgggtcgctgctgttgccactacttcg 300
QY 301 agcaagctcaaccccaaacatcccatccatccgaacgagcagcagatacgcacttcg 360
|||||
Db 301 agcaagctcaaccccaaacatcccatccatccgaacgagcagcagatacgcacttcg 360
QY 361 tctgttaaaccttgagctgctgctgctcccaattgcccacttagcaggttaacgcagcat 420
|||||
Db 361 tctgttaaaccttgagctgctgctgctcccaattgcccacttagcaggttaacgcagcat 420
QY 421 gaggatcgcatcacaccccgccgcaaacgagcagcagccacccactcgatttttcgcgcct 480
|||||
Db 421 gaggatcgcatcacaccccgccgcaaacgagcagcagccacccactcgatttttcgcgcct 480
QY 481 aagcggcaagagtcctcaaaccaacagcttcgagcagcaaacactcagcaagcagtcga 540
|||||
Db 481 aagcggcaagagtcctcaaaccaacagcttcgagcagcaaacactcagcaagcagtcga 540
QY 541 cccgagtgcactgttctcgagcagacacagaaagacgtcaacttcggacgcccga 600
|||||
Db 541 cccgagtgcactgttctcgagcagacacagaaagacgtcaacttcggacgcccga 600
QY 601 cagcacccgtcccaaatcccgagagcgcgcaagcccaagcagcagcagtcgaatccgc 660
|||||
Db 601 cagcacccgtcccaaatcccgagagcgcgcaagcccaagcagcagcagtcgaatccgc 660
QY 661 taattgatcagtgatgatactgctgctgctgctgctgctgctgctgctgctgctgctgct 720
|||||
Db 661 taattgatcagtgatgatactgctgctgctgctgctgctgctgctgctgctgctgctgct 720
QY 721 gcaagcaaccaatcaggaacagcctgatacagcagctccttccagaacacggcgagct 780
|||||
Db 721 gcaagcaaccaatcaggaacagcctgatacagcagctccttccagaacacggcgagct 780
QY 781 cgggtacacggcgagcagatgagcgggagcgggtacacggagtcgacagcagcgagcgg 840
|||||
Db 781 cgggtacacggcgagcagatgagcgggagcgggtacacggagtcgacagcagcgagcgg 840
QY 841 cgggtacacggcgagcagatgagcgggagcgggtacacggagtcgacagcagcgagcgg 900
|||||
Db 841 cgggtacacggcgagcagatgagcgggagcgggtacacggagtcgacagcagcgagcgg 900
QY 901 tggcggcagcgggtgagcggcgacacccactgcaacaggtgagcggcgagcggcgagcgg 960
|||||
Db 901 tggcggcagcgggtgagcggcgacacccactgcaacaggtgagcggcgagcggcgagcgg 960
QY 961 cactgtcaaacggcggtgagcgggtgagcgggtgagcgggtgagcgggtgagcgggtgagcgg 1020
|||||
Db 961 cactgtcaaacggcggtgagcgggtgagcgggtgagcgggtgagcgggtgagcgggtgagcgg 1020
QY 1021 ccttaacgctactcagtgatcgtgctgctgctgctgctgctgctgctgctgctgctgctgct 1080
|||||
Db 1021 ccttaacgctactcagtgatcgtgctgctgctgctgctgctgctgctgctgctgctgct 1080
QY 1081 cggcgaagatcaatgtgtgaaagacacatcaagagtcgagcgtgagcgtgagcgttgaagg 1140
|||||
Db 1081 cggcgaagatcaatgtgtgaaagacacatcaagagtcgagcgtgagcgttgaagg 1140
QY 1141 ccaagcgcgcaacttcaactgacgacgaacaaatcaatgaggttaagggagcagggcgaaatca 1200
|||||
Db 1141 ccaagcgcgcaacttcaactgacgacgaacaaatcaatgaggttaagggagcagggcgaaatca 1200
QY 1201 gaagcccatgttcgagctgctgagcggcgtacagttgaagaagtgaacctgggtgagaa 1260
|||||
Db 1201 gaagcccatgttcgagctgctgagcggcgtacagttgaagaagtgaacctgggtgagaa 1260
QY 1261 cgaagtcgatgtgcatccaagtgaaagccaaacgcctcaggaagtcacatvtgaacagt 1320
|||||
Db 1261 cgaagtcgatgtgcatccaagtgaaagccaaacgcctcaggaagtcacatvtgaacagt 1320

```
OY 1321 gcatgccagaacgctgcgtgaagacctgattacggtcacaaggcgagcgagcgacgcgct 1380
    |||||||
DB 1321 gcatgccagaacgctgcgtgaagacctgattacggtcacaaggcgagcgagcgacgcgct 1380
OY 1381 cactaatctgaacatcaagaacagcagctgcgaaggtgcagacgaacaaggtgtccagct 1440
    |||||||
DB 1381 cactaatctgaacatcaagaacagcagctgcgaaggtgcagacgaacaaggtgtccagct 1440
OY 1441 caagcgcaacatctggaatgaacactcaaggcgagcgagcgagcgagcgagcgagcgag 1500
    |||||||
DB 1441 caagcgcaacatctggaatgaacactcaaggcgagcgagcgagcgagcgagcgagcgag 1500
OY 1501 tcgcgcccaacgctgcgaagcagcttgatgacatgacgcgcgcgaagcgatcgaaac 1560
    |||||||
DB 1501 tcgcgcccaacgctgcgaagcagcttgatgacatgacgcgcgcgaagcgatcgaaac 1560
OY 1561 taaccacgagcaagctgcgcctgtgaaagcgacgctgacgcttgaagctgcgaacgag 1620
    |||||||
DB 1561 taaccacgagcaagctgcgcctgtgaaagcgacgctgacgcttgaagctgcgaacgag 1620
OY 1621 caacatcgcatgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1680
    |||||||
DB 1621 caacatcgcatgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1680
OY 1681 caccgagcttgatccagacagaagctggaataaagggtgagctc 1729
    |||||||
DB 1681 caccgagcttgatccagacagaagctggaataaagggtgagctc 1729
```

```
RESULT 3
AAAI4943
ID AAI4943 standard; DNA; 1729 BP.
XX
AC AAI4943:
XX
DT 08-AUG-2000 (first entry)
XX
DE DNA encoding a hypersensitive response elicitor protein.
XX
KW Hypersensitive response; insect control; disease resistance;
KW hypersensitive response elicitor; plant growth; vegetable; crop;
KW ornamental plant; ss.
XX
OS Pseudomonas syringae.
XX
FH Key Location/Qualifiers
FT CDS 419..1693
FT /tag= a
FT /product= "hypersensitive response elicitor protein"
XX
PN WO2000020452-A2.
XX
PD 13-APR-2000.
XX
PF 05-OCT-1999; 99WO-US23181.
XX
PR 05-OCT-1998; 98US-0103050.
XX
PA (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI Wei Z, Fan H, Niggemeyer JL;
XX
DR WPI: 2000-303745/26.
XX
DR P-PSDB: AAY84859.
XX
PT Hypersensitive response elicitor polypeptides useful for imparting
XX enhanced growth, disease resistance and insect resistance to plants,
XX especially vegetables and ornamental flowers -
XX
PS Disclosure: Page 25-26; 100pp; English.
XX
CC The present sequence encodes a hypersensitive response elicitor
CC polypeptide. The specification describes hypersensitive response
```

elicitor polypeptide fragments, which do not elicit a hypersensitive response. Instead, the proteins impart disease resistance to plants, enhance plant growth, and/or control insects. The polypeptide fragments may be used to these properties to plants. The plants which may be treated in this way include vegetables, crops and ornamental plants such as alfalfa, rice, wheat, barley, cotton, sunflower, peanut, corn, potato, sweet potato, bean, pea, chickory, lettuce, endive, cabbage, brussels sprout, beet, parsnip, turnip, cauliflower, broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon, citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco, tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia, petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.

Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 other.

Query Match 100.0%; Score 1729; DB 21; Length 1729;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 tccactgcgtgatttgaattgcaagatcattatagaacgttcaggtgtggaatcagg 60
    |||||||
DB 1 tccactgcgtgatttgaattgcaagatcattatagaacgttcaggtgtggaatcagg 60
OY 61 ctgagtcgcgagatttcgtgataagggtgtgtactgtaattgttgcattccaag 120
    |||||||
DB 61 ctgagtcgcgagatttcgtgataagggtgtgtactgtaattgttgcattccaag 120
OY 121 cctctgagtcggtgcgagacataccagcttctcgtcgtgcgtgtgcacactgagtcg 180
    |||||||
DB 121 cctctgagtcggtgcgagacataccagcttctcgtcgtgcgtgtgcacactgagtcg 180
OY 181 aggcataaggcatttcaagttcctctgctgtgttgagcattataaaggaaacttttaa 240
    |||||||
DB 181 aggcataaggcatttcaagttcctctgctgtgttgagcattataaaggaaacttttaa 240
OY 241 acagtcgaatgagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300
    |||||||
DB 241 acagtcgaatgagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300
OY 301 agcagctcaaccccaaacatccacatccctatcgaaacgagcagatcgacacttgc 360
    |||||||
DB 301 agcagctcaaccccaaacatccacatccctatcgaaacgagcagatcgacacttgc 360
OY 361 tctgttaaaccttgagctgcgtgcgttcacatttgcacattagcagagtaacgagcat 420
    |||||||
DB 361 tctgttaaaccttgagctgcgtgcgttcacatttgcacattagcagagtaacgagcat 420
OY 421 gagcatcgcatcacaccccgccgcgaacagacacacacgacacactcgatttcggcgt 480
    |||||||
DB 421 gagcatcgcatcacaccccgccgcgaacagacacacacacgacacactcgatttcggcgt 480
OY 481 aagcgcgaagagcttcctcaacacacacacacacacacacacacacacacacacacac 540
    |||||||
DB 481 aagcgcgaagagcttcctcaacacacacacacacacacacacacacacacacacacac 540
OY 541 aagcgcgaagagcttcctcaacacacacacacacacacacacacacacacacacacac 600
    |||||||
DB 541 aagcgcgaagagcttcctcaacacacacacacacacacacacacacacacacacacac 600
OY 541 cccgagtgacactgttctgcgagcagacacacacacacacacacacacacacacacac 660
    |||||||
DB 541 cccgagtgacactgttctgcgagcagacacacacacacacacacacacacacacacac 660
OY 601 cagcaccgtccagaatccgagagcagcagcagcagcagcagcagcagcagcagcagcag 660
    |||||||
DB 601 cagcaccgtccagaatccgagagcagcagcagcagcagcagcagcagcagcagcagcag 660
OY 661 taattgatcagtgatgatcatatgtcgttgcagatgctaccacactccaataaaaa 720
    |||||||
DB 661 taattgatcagtgatgatcatatgtcgttgcagatgctaccacactccaataaaaa 720
OY 721 gaagacacacacacacacacacacacacacacacacacacacacacacacacacacac 780
    |||||||
DB 721 gaagacacacacacacacacacacacacacacacacacacacacacacacacacacac 780
```

QY	761	cggtataacccgttcggtacgtatctgctggggcggtcggttataccggtatcgaaagttggtgcggtg	840
Db	761	cggtataacccgttcggtacgtatctgctggggcggtcggttataccggtatcgaaagttggtgcggtg	840
QY	841	cggtgatacgcgaagcgcaacacagtcggttgcggtcggttgaataccgcgcgaacacagcggtg	900
Db	841	cggtgatacgcgaagcgcaacacagtcggttgcggtcggttgaataccgcgcgaacacagcggtg	900
QY	901	tgggtggcagcggttggcggcggtcacacacacatctgcgaacagttggtgcggtcgaacacgc	960
Db	901	tgggtggcagcggttggcggcggtcacacacacatctgcgaacagttggtgcggtcgaacacgc	960
QY	961	cactgcgaacagcggttggcgggttggcgttgaacaccgcgaatactcctgcgcagttggccaa	1020
Db	961	cactgcgaacagcggttggcgggttggcgttgaacaccgcgaatactcctgcgcagttggccaa	1020
QY	1021	cccttaaccggtatacctcaaggtatctgcgttcgttcggaacacgcgcgaagttcttaaccggaagc	1080
Db	1021	cccttaaccggtatacctcaaggtatctgcgttcgttcggaacacgcgcgaagttcttaaccggaagc	1080
QY	1081	cggtcaagatcaatgtgtgtgaagaacacacatcaaggttcggcgttcgggtcggaagtttgacgtg	1140
Db	1081	cggtcaagatcaatgtgtgtgaagaacacacatcaaggttcggcgttcgggtcggaagtttgacgtg	1140
QY	1141	ccacggtgcgaacctctcactctgcgcgaacaatctatgtgtgaacggtgagacacaggtcgaaataca	1200
Db	1141	ccacggtgcgaacctctcactctgcgcgaacaatctatgtgtgaacggtgagacacaggtcgaaataca	1200
QY	1201	gaagcccatgtttgcgagcttggtcttgaagcgctctacgtttgaagaaatgtgaaacctgggttgagaa	1260
Db	1201	gaagcccatgtttgcgagcttggtcttgaagcgctctacgtttgaagaaatgtgaaacctgggttgagaa	1260
QY	1261	cgaggttcgatgtgcataccacgttggaagcgaacaaacgcgtcgaagatcaccatcttgacaagct	1320
Db	1261	cgaggttcgatgtgcataccacgttggaagcgaacaaacgcgtcgaagatcaccatcttgacaagct	1320
QY	1321	gcatgtcccaagaaacgttcggttgaagaacctgtatcaagttcaaaagcggaagcggtgcgcagctg	1380
Db	1321	gcatgtcccaagaaacgttcggttgaagaacctgtatcaagttcaaaagcggaagcggtgcgcagctg	1380
QY	1381	cactcaatctgtaaatctcaagatacagacaggtgcgaaggtgcgaacaggtgtgtccagct	1440
Db	1381	cactcaatctgtaaatctcaagatacagacaggtgcgaaggtgcgaacaggtgtgtccagct	1440
QY	1441	caagcgcaaacacacacacttggaatactgaacacatctcaagcggtcgaacgtttcttcggtcagatggt	1500
Db	1441	caagcgcaaacacacacacttggaatactgaacacatctcaagcggtcgaacgtttcttcggtcagatggt	1500
QY	1501	tgcgaccaaagcttgcgacaaagctttgtatgacatgtgacatcgagctcggtgaacggtcatcgaaagc	1560
Db	1501	tgcgaccaaagcttgcgacaaagctttgtatgacatgtgacatcgagctcggtgaacggtcatcgaaagc	1560
QY	1561	taacccaagcggaagttgcgacctgtgttgaaaaagcggaacgttgaacgtctgaaagcttgcgaacggtg	1620
Db	1561	taacccaagcggaagttgcgacctgtgttgaaaaagcggaacgttgaacgtctgaaagcttgcgaacggtg	1620
QY	1621	caaatctgcacatctacccaagcttcaacaacgcttaccgatataaaccgaagatctgacacaca	1680
Db	1621	caaatctgcacatctacccaagcttcaacaacgcttaccgatataaaccgaagatctgacacaca	1680
QY	1681	caccgagctttgaatccagacaagtagcttgaataaaaggggttgtagctc	1729
Db	1681	caccgagctttgaatccagacaagtagcttgaataaaaggggttgtagctc	1729
RESULT 4			
AAK09007 standard; DNA; 1344 BP.			
XX	AC	AAK09007;	
XX	TT	14-JUN-1999 (first entry)	

DE	Hypersensitive response eliciting protein coding sequence (hrpM).	
XX		
KW	Hypersensitive response elicitor protein; hypersensitive response;	
KW	hrpM; pathogen; infection; crop protection; disease resistance;	
KW	pest resistance; transgenic plant; colouration; maturation; ss.	
OS	Erwinia amylovora.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1344
FT		/*tag= a
FT		/product= Hypersensitive_response_elicitor_protein
XX		
PN	WO9907208-A1.	
XX		
PD	18-FEB-1999.	
XX		
PF	27-JUL-1998;	98WO-US15547.
XX		
PR	06-AUG-1997;	97US-005108.
XX		
PA	(CORR) CORNELL RES FOUND INC.	
XX		
PI	Beer SV, Kim JF;	
XX		
DR	WPI: 1999-167126/14.	
DR	P-PSDB: AAW96260.	
XX		
FT	New Erwinia amylovora hypersensitive response eliciting gene and	
FT	protein - useful for providing transgenic plants and seeds with	
PT	enhanced growth, and insect and disease resistance	
XX		
PS	Claim 1; Page 49; 54pp; English.	
XX		
CC	The hypersensitive response eliciting protein (hrp) or polypeptide	
CC	is produced as part of an active defense by plants against	
CC	incompatible pathogen infections. The hypersensitive response is a	
CC	rapid localised necrosis. The hrp protein and gene when used in	
CC	nucleotide constructs are useful for providing disease resistance to	
CC	plants, insect control to plants, and enhancing plant growth	
CC	(enhancing fruit size and earlier colouration and maturation), by	
CC	direct application of the protein to plants, or by producing	
CC	transgenic plants or seeds using the hrp gene.	
XX		
Sequence 1344 BP; 343 A; 372 C; 357 G; 272 T; 0 other;		

Query Match	Score	DB20	Length
Best Local Similarity	57.18	Pred. No. 8, 3e-30	
Matches	330	Conservative	0; Mismatches 233; Indels 15; Gaps
QY 1079	gcccgcgaatcatcatgtgtggtgaagaacacatcatcaagtcgcgcgtcgcgaagctttgac	1138	
Db 745	ggcgctaatcagaagcggtgcgtcatgaacattaccgtgaagcgggtcagtggttat	804	
QY 1139	ggccacggcgcaaccttcaatgccgcgaacaattcatgtgttaacggaagaccagggcgaaat	1199	
Db 805	ggcacaagaaacaaacttaccgcgcggtttcgaattggcgatggcggccaagtctyaaaac	864	
QY 1199	cagaaagccatgttcgcagctgcgtctgaagcggtcagctgtgaagaattgtaacctgtgtgaa	1256	
Db 865	cagaaacccgtgtttattaccgtgaagacggtgtgcagctgtgaaaaacgtcaccttggcgac	924	
QY 1259	aacgaagtgcagtgcacatcagctgtgaagccaaaacgctcaggaagttacatctgacaac	1318	
Db 925	gacggggtcgagatggtattcactcttcacg-----gtgatggccaaatagaacaat	972	
QY 1319	gtgatgtcccaagaagctcggttgaagaccgcgattacggtgaagaaggcgagggtgcgcagcg	1378	
Db 973	ctcgacgtcccaacagcggtgtgagagcgcgattacggttaagccaaacgcgcgggacaa	10322	
QY 1379	gtcactatctgaacttcaagcaaacacagctgtccaaagggtgcagacgacaaggttttcag	1438	

Db 1033 aaatccacgcttgtaatcaactaacagcttcttcctgcagcaagcgcctctgcataacagatcctgag 10922
 QY 1439 ctcaacgccaacactcaacttggaaatcgcgaacacttcaagcgccgaagatlttgcgcagatg 1498
 Db 1093 ctgaatgcgatactcaactcgcgagcgtttgcgaacacgtaagacccaagaagacttgcatactt 11522
 QY 1499 gttcgacccaacagcgtgcgaacgcggtttgtatgtacatgagcatcgaaactgaacgcgcgaa 1558
 Db 1153 gtaacgacactaacgcggtcaacag---ggtaactcgtgatactgcatactgcgcataatcgc 1209
 QY 1559 gctaacacagcgcaagttcgcctcgtgtgaaagcgacgtgcacatctcgaagcctgcgaacg 1618
 Db 1210 gcacaaagacgcgtatgctctcgttcgtttaaagcgcataagcgagggcgtataacgctaatcc 1269
 QY 1619 ggcacacgtcgcaatccacgcgtcaacacgcgcctacga 1656
 Db 1270 agtgatactccactcgtgtagatgttgcgaacacactaa 1307

RESULT	5
AAD00669	
ID	AAD00669 standard; DNA; 1344 BP.

AC AAD00669;

DT 08-SEP-2000 (first entry)

DE Erwinia amylovora hypersensitive response elicitor encoding DNA #2

KW Hypersensitive response elicitor; environmental stress resistance;

KW plant; ds.

OS Erwinia amylovora.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

F ^T	CDS	1..1344
----------------	-----	---------

FT /product= "Hypersensitive response elicitor"

PN WO200028055-A2.

PD 18-MAY-2000

PF 04-NOV-1999; 99WO-US26039.

PR 05-NOV-1998; 98US-0107243.

PA (EDEN-) EDEN BIOSCIENCE CORP.

PI Wei Z, Schading RL;

DR WPI; 2000-376566/32.

XX.

PT impact stress resistance -

PS Disclosure; Page 10; 84pp; English.

The pathogenesis of a method to impart stress resistance to plants by CC applying a hypersensitive response elicitor in a non-infectious form to CC a plant or seed. The present sequence is a DNA encoding hypersensitive CC response elicitor protein from *Erwinia amylovora*. The protein is CC heat stable, protease sensitive and suppressed by inhibitors of plant CC metabolism. The present sequence is used to transform transgenic plant CC or plant seeds to impart stress resistance.

SD Sequence 1344 BP; 343 A; 372 C; 357 G; 272 T; 0 other;

Query Match	9.0%;	Score 155.2;	DB 21;	Length 1344;
Best Local Similarity	57.1%;	Pred. No. 8.3e-30;		
Matches 330;	Conservative 0;	Mismatches 233;	Indels 15;	Gaps 2

QY	1079	gcccgcgaagttcaatctgtgtgaaagaccacattcaaggttcggtgcgtgcgaagctttcac	1138
Db	745	ggcgcttaacagacggtgtgcgtgcatactgacacattacccgtgaaagcggtgcaggtgtttgat	804
QY	1139	ggccacagcgcaacaccttcaactgctgcgcgaacaattatgtgttaacgtagacccagcgcyanaat	1198
Db	805	ggccaaagacaaaccttcaacgcgcgtgttcagatttagtgatgtgcgtgcgtctctaaac	864
QY	1199	cagaagcccatgtttcgaagctgtcgtctgtgaagcgctgaagcttgaagaatgttaacctgtgtag	1258
Db	865	cagaaacccgctgttttaactgtgaagcagcggtgcgcacgttcaaaacgltacatgtgtgcac	924
QY	1259	aacgaagttgattgtgcctccacgtagaagccaaaacgctcaggaagtacatttcacac	1318
Db	925	gaacggcggtgattgtgtatctcaactttagc-----gtgagccaaatagacat	972
QY	1319	gtgcattccccagaagcgttcggtgtgaagacactgtattacgltcaaaagcgagggagcgccagc	1378
Db	973	ctgtagcttaaccacaaactgtggtgtgagcgcggtattaccgtttaagccaacacgctgcgcaca	1032
QY	1379	gtcaactaacttgacaatcaaagacagcagttgccaaagttgcagacgcacaagttgttcag	1438
Db	1033	aaatcccaacgttaaatcaactaacagttctcttcgcagcaagcctctgcacaagaatccttgag	1092
QY	1439	ctcaagcccaacacctcaactgaanaatcgcgaacattcaagcgccagcagatttcgcgcagtg	1498
Db	1093	ctgaatgtcatcttaacctgaacctggtttgcacaacgtgaagggccaaagaactttgttaacttt	1152
QY	1499	gttcgcaccacaagcgttgcacaagcagtttctgtatgtgacatgagctgaacgycatcgaa	1558
Db	1153	gttaagcactaagcgcgtgtcaacag---ggttaactgttgatctgatatctgagccatacagc	1209
QY	1559	gctaacacaggtcaggttgcgtccctgtgtgaaagcgcagatgtacgactctgaagctgtgcacg	1618
Db	1210	gcagaagacgctgaagttctcgtcttcgtttaaaagcgcagatagcgaggtgcctaactcatacc	1269
QY	1619	ggcacaatcgcatagcagcgcgtgcacaagcgcttaaga	1656
Db	1270	agtgtatacttcactgtgtgtatgtttgaaacccactaaa	1307

RESULT 6

ID	AAA14939	standard; DNA; 1344 BP.
----	----------	-------------------------

AC AAA14939;

DT 08-AUG-2000 (first entry)

DE DNA encoding a hypersensitive response elicitor protein.

KW Hypersensitive response; insect control; disease resistance;

KW ornamental plant; ss.

OS Erwinia amylovora

FH	Key	Location/Qualifiers
1	334	

```
FT      /*tag= a "XXXXXXXXXXXXX" proto="P
```

PN W0200020452-A2

PD 13-APR-2000.

PF 05-OCT-1999; 99WO-US23181.

PR 05-OCT-1998; 98US-0103050.

PA (EDEN-) EDEN BIOSCIENCE CORP

XX

Db 10564 nvaasrshrnsry 10576

RESULT 9

AAQ21833 ID

AAQ21833 standard; DNA; 390 BP.

XX

AC AAQ21833;

XX

DT 08-JUN-1992 (first entry)

XX

DE Randomising oligonucleotide used in SPERT mRNA prep.

XX

KM Systematic polypeptide evolution by reverse translation; SPERT;

XX ligand binding; ss.

XX

OS Synthetic.

XX

PN WO9202536-A.

XX

PD 20-FEB-1992.

XX

PF 01-AUG-1991; 91WO-US05463.

XX

PR 02-AUG-1990; 90US-0561968.

XX

XX (COLS) UNIV OF COLORADO.

PA

XX Gold L, Tuerk C;

PI

XX WPI; 1992-080018/10.

DR

XX

PT New method of systematic polypeptide evolution by reverse

PT translation - by linking each polypeptide in sample mixt. to

PT individualised mRNA allowing further synthesis of selected

PT polypeptide(s)

XX

XX

PS Example; Page 55; 102pp; English.

XX

CC The sequence is that of an example randomising oligonucleotide which

CC is used in the prep. of mRNA encoding candidate polypeptides for the

CC method of systematic polypeptide evolution by reverse translation

CC (SPERT). The method provides a rapid way of isolating and identifying

CC polypeptide ligands which bind to target mols. The polypeptide ligands

CC can be used in e.g. assay methods, diagnostic procedures, cell sorting,

CC as activators or inhibitors of target mol. function, as probes, as

CC sequestering agents, drug delivery vehicles, modifiers of hormone

CC action and as catalysts. See also AAQ21830-Q21832.

CC

XX

XX Sequence 390 BP; 125 A; 126 C; 133 G; 6 T; 0 other;

SQ

Query Match 3.3%; Score 56.6; DB 13; Length 390;

Best Local Similarity 47.1%; Pred. No. 5e-05;

Matches 173; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 1256 gagaacgagtgatgcatccacagtgaaagccaacacgtcagagaccattgac 1315

DB 21 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 80

QY 1316 aacgtcatccccaagacgctggtgaagacctgattcgttcaaaagcgaggagcgca 1375

DB 81 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 140

QY 1376 gcggtcaactatgtgacatcaagacagcagtgccaaggtgcagaacaggtgtc 1435

DB 141 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 200

QY 1436 cagctcaacgccaacactccttgaatgacaacttcaagccgagcttcggcagc 1495

DB 201 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 260

QY 1496 atggttcgcacccaagcgtgcagaagcttgatgacatgacgtcagtcgaacgcgcatc 1555

Db 261 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 320

QY 1556 gaagctaacccaagcagcttcgacctgltgaaagcagcagltgacgactgaagctgcga 1615

DB 321 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 380

QY 1616 acggcgca 1622

DB 381 atggtca 387

RESULT 10

AAQ36859 ID

AAQ36859 standard; DNA; 390 BP.

XX

AC AAQ36859;

XX

DT 22-JUN-1993 (first entry)

XX

DE PCR primer for 5' fixed sequence contg. T7 promoter and RBS.

XX

KM Systematic peptide evolution by reverse translation; SPERT; ligand;

XX specific; inhibitors; probes; assay; cell sorting; ss.

XX

OS Synthetic.

XX

PN WO9303172-A.

XX

PD 18-FEB-1993.

XX

PF 31-JAN-1992; 92WO-US00801.

XX

PR 01-AUG-1991; 91US-0739055.

XX

XX (UYRE-) UNIV RES CORP.

PA

XX

PI Gold L, Prihnow D, Smith JD, Tuerk C;

XX

DR WPI; 1993-076529/09.

XX

PT Systematic polypeptide evolution by reverse translation - used

PT for prod. of polypeptide ligand specific for desired target

PT molecule

XX

XX

PS Example 1; Page 84; 98pp; English.

XX

CC SPERT is used to select novel polypeptides that bind the antibody

CC of the epitope commonly recognised by the antisera from autoimmune

CC mice which are the F1 progeny of a cross of NZB and NZW parents

CC (Portanova et al., J. Immunol. 144, 4633, 1990). The known epitope

CC consists of ca. 10 amino acids at the N-terminus of the histone H2B

CC protein. To make mRNA encoding candidate polypeptides a 5' fixed

CC sequence composed of a T7 promoter sequence and a ribosome binding

CC site which is recognised by both prokaryotic and eukaryotic ribosomes,

CC terminating in a restriction endonuclease site is synthesised and cloned

CC using a number of oligonucleotides (example shown). A 3' fixed sequence

CC is placed into a restriction site to provide an mRNA encoding the C-

CC terminal trailer sequence of ca. 100 nucleotides lacking stop codons.

CC In addition, a 3' primer annealing site is provided so that cDNA

CC synthesis can be accomplished on the mRNA recovered from partitioned

CC ribosome complexes. See also AAQ36845-63.

XX

XX Sequence 390 BP; 125 A; 126 C; 133 G; 6 T; 0 other;

SQ

Query Match 3.3%; Score 56.6; DB 14; Length 390;

Best Local Similarity 47.1%; Pred. No. 5e-05;

Matches 173; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 1256 gagaacgagtgatgcatccacagtgaaagccaacacgtcagagaccattgac 1315

DB 21 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 80

```
QY 1336 aacgtgcatgccagaacgtcggtgaagacctgattacggtcaaaagcgaggagcgca 1375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 140
QY 1376 ggggtcactaatctgaactcaagaacgacgacgtgcgaaggtgcagacgaagttgtc 1435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 200
QY 1436 gacgtcaacgcaacactactctgaaatcgaactcaagtcgacgacgttcgacgacg 1495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 260
QY 1496 atggttcgacccaacggtggaagcagtttgatgacatgacatgacgtgaacgacac 1555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 320
QY 1556 gaagctaacccagcgcaagtcgacctgtgtgaaagcgacagtcgacatctgaagctgca 1615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 380
QY 1616 acgggca 1622
    ||| ||| |||
Db 381 atgttca 387
```

RESULT 11

AAAF6910 standard; DNA; 390 BP.

AAAF6910;

29-MAY-2001 (first entry)

Sequence containing a 120 repeat of ACG flanked by fixed fragments.

Ligand isolation; systemic polypeptide evolution by reverse translation;

SPERT; ss.

Synthetic.

US6194550-B1.

27-FEB-2001.

23-NOV-1998; 98US-0197649.

31-JAN-1992; 92US-0829461.

02-AUG-1990; 90US-0561968.

01-AUG-1991; 91US-0739055.

(GOLD/) GOLD L.

(TUBER/) TUBER C.

(PRIB/) PRIBNOW D.

(SMIT/) SMITH J D.

Gold L, Tuerk C, Prihnow D, Smith JD;

WPI; 2001-243412/25.

Isolating a polypeptide ligand to a target molecule, useful for

diagnostic assays, comprises partitioning candidate mixtures comprised

of ribosome complexes or mRNA, polypeptide copolymers relative to their

affinity to the target molecule

Example; Column 39; 35pp; English.

The present sequence was used in an example illustrating an invention

relating to a method for isolating a polypeptide ligand for a desired

target molecule. The method involves synthesizing a nucleic acid mixture

comprising mRNA having translatable and non-translatable regions and a

mixture of nucleic acid, polypeptide copolymers, each comprising the mRNA

and a polypeptide encoded by its associated mRNA. The copolymers are

partitioned relative to their affinity to the target. The method is
termed systemic polypeptide evolution by reverse translation (SPERT).
The polypeptides ligands of small molecule targets are useful in assay
methods, diagnostic procedures, cell sorting, as inhibitors of target
molecule function, as probes, as drug delivery vehicles and modifiers of
hormone action and have therapeutic uses as sequestering agents. The
target molecules include natural and synthetic polymers, including
proteins, hormones, receptors and cell surfaces, nucleic acids and small
molecules such as drugs, metabolites, cofactors and toxins. Polypeptide
ligands are isolated and rapidly identified by this method.

Sequence 390 BP; 125 A; 126 C; 133 G; 6 T; 0 other;

Query Match 3.3%; Score 56.6; DB 22; Length 390;
Best Local Similarity 47.1%; Pred. No. 5e-05;
Matches 173; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 1256 ggaagcaggtcgatgcatccacgtgaaagccaagcctcaggaagtcacacattgac 1315

Db 21 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 80

QY 1316 aacgtgcatgccagaacgtcggtgaagacctgattacggtcaaaagcgaggagcgca 1375

Db 81 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 140

QY 1376 ggggtcactaatctgaactcaagaacgacgacgtgcgaaggtgcagacgaagttgtc 1435

Db 141 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 200

QY 1436 gacgtcaacgcaacactactctgaaatcgaactcaagtcgacgacgttcgacgacg 1495

Db 201 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 260

QY 1496 atggttcgacccaacggtggaagcagtttgatgacatgacatgacgtgaacgacac 1555

Db 261 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 320

QY 1556 gaagctaacccagcgcaagtcgacctgtgtgaaagcgacagtcgacatctgaagctgca 1615

Db 321 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 380

QY 1616 acgggca 1622

Db 381 atgttca 387

RESULT 12

AAAF69879 standard; DNA; 591 BP.

AAAF69879;

28-JAN-1999 (first entry)

Bacillus sp strain KSM-P15 pectic acid lyase encoding DNA.

Bacillus sp. strain KSM-P15; pectic acid lyase; proteopeptinase;

proteopeptin; polygalacturonic acid; detergent; cotton; surfactant;

cellulase; protease; bleaching agent; ds.

Bacillus sp.

WO9845393-A2.

15-OCT-1998.

08-APR-1998; 98WO-JP01613.

Location/Qualifiers

1..591

/*tag= a

Key CDS

```

PR 08-SEP-1997; 97JP-0242736.
PR 09-APR-1997; 97JP-0091142.
XX (KAOS ) KAO CORP.
XX
XX Hataeda Y, Ito S, Kasai M, Kobayashi T, Koike K;
XX Shikata S, Suzumatsu A, Tsumadoni M, Wada Y;
XX WPI: 1998-568339/48.
XX P-PSDB: AAM83014.
XX
XX Detergent composition containing protopectinase active at alkaline
XX pH - on protopectin and polygalacturonic acid, provides better
XX removal of muddy soil
XX
XX Example: Page 71-72; 80pp; English.
XX
XX The present invention describes a detergent composition which contains
XX a protopectinase having an optimum pH 7 or higher against protopectin
XX and polygalacturonic acid substrates. Inclusion of protopectinase
XX gives a composition that provides better removal of muddy soil,
XX particularly from socks. The present sequence encodes pectic acid lyase
XX from Bacillus sp. strain KSM-P15, which is used in an example from
XX the present invention. Pectic acid lyase exhibits protopectinase
XX activity.
XX
XX Sequence 591 BP; 160 A; 134 C; 179 G; 118 T; 0 other;
XX
XX
XX Query Match 3.3%; Score 56.6; DB 19; Length 591;
XX Best Local Similarity 50.8%; Pred. No. 5.9e-05;
XX Matches 236; Conservative 0; Mismatches 199; Indels 30; Gaps 3
XX
OY 1094 gtgtggaagaccatcacgaagtcgctgcgtgcgaagctttgacggcagcgcaacc 1153
DB 10 gtgcgtcatgaacgacattctgtgtcgtccggtcagacttgaagcgaagaagcgacc 69
OY 1154 t---tcactgcgcgaacaatctatgggtaaaggagaccaggcgcaaatcagaagccatg 1210
DB 70 tatgtggttaatccgaatcatcttggggagcgatcgcagcgagatcagaagccgacc 129
OY 1211 ttcgagctgcgcgaagcgctgactgttgaagaatgtgaacattggtgtgagaagagtcgat 1270
DB 130 ttctgcctggagagctcggggcgaagccttgaaaaatgtagatgtgcctctcgcgtgac 189
OY 1271 ggcataccagtgtaaaagccaaaaacgctcaggaagtcacattgacacgtgatgcgacg 1330
DB 190 ggggtgcactgctaag-----ggattgtaagattacaatgatcatcttgagag 237
OY 1331 aacgtcgtgtgaagacctgattcaggtcgaaggcgagggagcgacggtcactaatctg 1390
DB 238 gatgtctgtgtgaagatcgtgctgacgcttaa-----atcgtccggaaagtggt 282
OY 1391 aacatcgaagacagcagctgtccaaagttcgagacgcaagagtttccaggtcgaagcgcaac 1450
DB 263 aacatctcggggcgggcgacgctcaacaggcgatgtgacaagtggttccaatcaatgacg 342
OY 1451 actcactggaataatcgacaacattcaagcgcgacgatttcgacagatgtgtgcaccaac 1510
DB 343 gggagcatcaaacattcgttaacttcagggcgatgcatcgggaagtggttctcgcaaac 402
OY 1511 ggtgcgaagcaggttgatgacacttgagcatctgcggtcgaagcgatc 1555
DB 403 ggaagcaccacactcacaagtggtgtgaagcgtggaataaactcgcaac 447
XX
XX RESULT 13
XX AAX89484
XX ID AAX89484 standard; DNA; 666 BP.
XX AC AAX89484;
XX
XX 17-JAN-2000 (first entry)

```

	XZ	Bacillus licheniformis DNA encoding pectate lyase I.
DE	XZ	Pectate lyase I; EC 4.2.2.2; pectin degrading enzyme; ATCC 14580;
KM	XZ	catalytic active domain; cellulose binding domain; CBD; operably linked;
KW	XZ	optimum activity; pH: detergent composition; yarn; cellulosic fibre;
KM	XZ	recycled waste paper; pulp; retting process; animal feed; juice;
KX	XZ	transgenic plant; ss.
OS	XZ	Bacillus licheniformis.
FH	XZ	Key Location/Qualifiers
FT	CDS	/location= "1..666"
FT	a	/product= "Pectate lyase I"
FT	b	/EC_number= "4.2.2.2"
FT	c	/number= "82..663"
FT	d	/function= "Pectin degradation"
FT	e	/label= "Pro-sequence"
XX	XX	WO9927083-A1.
PN	XX	03-JUN-1999.
PD	XX	24-NOV-1998:
PE	XX	98MO-DK00514.
PR	XX	24-NOV-1997:
PR	XX	06-MAY-1998:
PA	XX	(NOVO) NOVO-NORDISK AS.
PI	XX	Andersen LN, Bjornvad ME, Lange NEK, Schnorr K, Schuelein M;
DR	XX	WPt: 1999-610578/52.
PS	XX	P-PDB: AAT28446.
PT	XX	New isolated pectate lyase enzymes -
PS	XX	Claim 7: Page 79; 93pp: English.
CC	XX	The present sequence is a DNA encoding pectate lyase I. It is a pectin
CC	XX	degrading enzyme derived from Bacillus licheniformis, ATCC 14580. The
CC	XX	enzyme comprises a catalytic active domain and a cellulose binding domain
CC	XX	(CBD) that are operably linked to each other. It shows optimum activity
CC	XX	at pH greater than 9 and temperature 55 degree centigrade. It can be used
CC	XX	In detergent compositions, for cleaning hard surfaces, for machine
CC	XX	treatment of fabrics, for improving the properties of cellulosic fibres,
CC	XX	yarn, woven or non-woven fabric, for the degradation of plant material
CC	XX	e.g., recycled waste paper, mechanical paper-making pulps or fibres
CC	XX	subjected to retting process, for preparing animal feed and for
CC	XX	processing wine or juice. DNA encoding the enzyme can also be used for
CC	XX	the production of transgenic plants.
SQ	XX	Sequence 666 BP; 205 A; 136 C; 178 G; 147 T; 0 other;
OY	Query Match	3.2%; Score 55.8; DB 20; Length 666;
Db	Best Local Similarity	52.6%; Pred. No. 9.9e-05;
Matches	Matches 154; Conservative	0; Mismatches 127; Indels 12; Gaps 1
OY	1090 caatygtgtaaaagcaccataaatgaagtgcggcgctgcgagaatttttagccgccaggcgc	1149
Db	87 cgaagtcgtttcaaaaacgatctgtagtcgagaagccccaaactgatatgcgaaaaaggcaa	146
OY	1150 aacctaacctgcgcgaacaatcataatggttgtaacgagaccagggcgaaaatcgaagcccat	1209
Db	147 ggcgcgcatgtgcagagctcgcgagactcggggagacgcccgaacgcyagatcaaaacgat	206
OY	1210 gtctcagatctgcgtgaagcgcgtacgtlttgaaagtgaacctgggtgtagaagcaggtcga	1269

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2001, 20:31:10 ; Search time 71.4 Seconds
(without alignments)
4584.299 Million cell updates/sec

Title: US-09-597-513-1
Perfect score: 1729
Sequence: 1 tccacttcgcgcgattttgaa.....tgaataaagggtggaactc 1729

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
4: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
5: /cgn2_6/prodata/2/lna/PCrUS_COMB.seq:*
6: /cgn2_6/prodata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1729	100.0	1729	4	US-09-120-817-1 Sequence 1, Appl1
2	56.6	3.3	390	4	US-09-197-649-7 Sequence 7, Appl1
3	56.6	3.3	591	4	US-09-402-668-1 Sequence 1, Appl1
4	55.8	3.2	666	4	US-09-198-956-3 Sequence 3, Appl1
5	53.6	3.1	7218	1	US-08-332-463-14 Sequence 14, Appl1
6	49.8	2.9	6530	2	US-08-146-930-1 Sequence 1, Appl1
7	49.8	2.9	6530	5	US-08-458-240-1 Sequence 1, Appl1
8	49.8	2.9	6530	5	US-08-458-240-1 Sequence 1, Appl1
9	47.4	2.7	5036	4	US-09-177-349-2 Sequence 2, Appl1
10	47.2	2.7	1280	4	US-09-060-756-4 Sequence 4, Appl1
11	47	2.7	1028	4	US-08-118-200-1 Sequence 1, Appl1
12	47	2.7	1028	4	US-08-118-200-1 Sequence 1, Appl1
13	45.8	2.6	1548	2	US-08-762-106-5 Sequence 5, Appl1
14	45.8	2.6	1548	2	US-08-762-106-5 Sequence 5, Appl1
15	44.6	2.5	68750	3	US-09-335-409-1 Sequence 6, Appl1
16	43.4	2.5	2793	1	US-08-209-747-1 Sequence 1, Appl1
17	43.4	2.5	2793	1	US-08-209-747-1 Sequence 1, Appl1
18	43.4	2.5	4378	2	US-09-080-897-3 Sequence 3, Appl1
19	43.4	2.5	4378	4	US-09-323-735-3 Sequence 3, Appl1
20	43.4	2.5	4399	3	US-08-899-955-2 Sequence 2, Appl1
21	43.2	2.5	12568	2	US-08-387-942C-1 Sequence 1, Appl1
22	42.6	2.5	1284	4	US-09-312-038-4 Sequence 4, Appl1
23	42.6	2.5	1284	4	US-09-312-038-4 Sequence 4, Appl1
24	42	2.4	185	4	US-09-402-668-9 Sequence 9, Appl1
25	42	2.4	530	3	US-08-758-662-4 Sequence 4, Appl1
26	42	2.4	1162	2	US-08-726-306A-52 Sequence 52, Appl1
27	42	2.4	4524	2	US-08-845-998-7 Sequence 7, Appl1

28	42	2.4	4524	3	US-09-206-537-7 Sequence 7, Appl1
29	41.8	2.4	1931	2	US-09-130-114-2 Sequence 2, Appl1
30	41.6	2.4	219	1	US-07-609-716-34 Sequence 34, Appl1
31	41.6	2.4	219	1	US-08-175-135-51 Sequence 51, Appl1
32	41.6	2.4	219	1	US-08-477-509B-86 Sequence 86, Appl1
33	41.6	2.4	219	2	US-08-477-509B-86 Sequence 86, Appl1
34	41.6	2.4	219	3	US-08-482-085B-86 Sequence 86, Appl1
35	41.6	2.4	219	3	US-08-475-411A-34 Sequence 34, Appl1
36	41.6	2.4	219	4	US-08-478-029A-34 Sequence 34, Appl1
37	41.4	2.4	18994	2	US-08-459-586-4 Sequence 4, Appl1
38	41.4	2.4	18994	2	US-08-282-696-4 Sequence 4, Appl1
39	41.2	2.4	1035	1	US-08-891-254-8 Sequence 8, Appl1
40	41.2	2.4	1035	2	US-08-819-539-8 Sequence 8, Appl1
41	41.2	2.4	1035	2	US-09-030-270A-8 Sequence 8, Appl1
42	41.2	2.4	1035	2	US-08-984-207-8 Sequence 8, Appl1
43	41.2	2.4	1035	5	US-08-984-207-8 Sequence 8, Appl1
44	40.6	2.3	1729	4	US-09-120-817-1 Sequence 1, Appl1
45	40.4	2.3	2505	1	US-07-977-434-7 Sequence 7, Appl1

ALIGNMENTS

RESULT 1
US-09-120-817-1
; Sequence 1, Application US/09120817
; Patent No. 6172184
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Charkowski, Amy
; APPLICANT: Alfano, James R.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: PSEUDOMONAS STRINGAE AND ITS USE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,817
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,107
; FILING DATE: 06-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; TELEPHONE/DOCKET NUMBER: 19603/1741
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1729 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-120-817-1
Query Match 100.0%; Score 1729; DB 4; Length 1729;
Best local Similarity 100.0%; Pred. No. 0;
Matches 1729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

Oy 1150 aaccttactgacgcacaatctatgtgtacaggaagcagccagggcgaaatccagaacccat 1203
Db 147 ggcgcgtcatgcaaggtccgcagagcttcggagcgagcgcaacgacgaggaatcaaaaacgcgt 206
Oy 1210 gttcagacgtgctbaagcgctactacgttgaagaatgtgaacctgggtgtgaacagagtcga 1268
Db 207 ttctaaagttgaggaatgtgtgcaacgctcacaataatgtcgtgtcttggtgcctccgcgtccga 266
Oy 1270 tggcatccacgttgaagccaaaacgcgtcagaagtcaccatgtgacaactgtacgtccca 13228
Db 267 tgggtttcacacacatattggaac-----gcttcataaacaacgttgttggga 314
Oy 1330 gaacgttcgggtgaagacctgatttcggttcaaaaggcgagggagcgcaaggtca 1382
Db 315 agatgttcggcgaagaatgtccttgactgtcacaagcgaaaggaagtgcacgataa 367

RESULT 5
US-08-232-463-14/c
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ. ID NO.: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgpt-Fls
: US-08-232-463-14

Query Match 3.1%; Score 53.6; DB 1; Length 7218;
Best Local Similarity 5.08; Pred. No. 0.00024;
Matches 20; Conservative 217; Mismatches 161; Indels 0; Gaps
1214 gagctggtgaagcgacctacgttgaagaatgtgaacctgtgtgagaaacgagtcgatgac 1273

```

[illegible]

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-146-930-1

Query Match 2.9%; Score 49.8; DB 2; Length 6530;
Best Local Similarity 50.6%; Pred. No. 0.0022;
Matches 120; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 755 gcccttcacagacacagcgagcggtcgtgacacgcgtcgcgcgacgagcgagcgagcggt 814
DB 3774 GCCCCCCAGAGAGCTACGAGAGGGGCTCTCCGAGAGAGGTGACTCTGAGAGTGGC 3833
QY 815 acacgcgagatcgacaggtgagcgagcggtgacacgcgacgacgagcgagcgagcg 874
DB 3834 TCCTCTGGCGGCGGTGGCGGCGGTGCTCTACTCCAGCGGTGTGGGCGGAGCAGCGGT 3893
QY 875 ggtgatactccgacgcgacagcggtgagcgagcggtgagcgagcggtgagcgagcg 934
DB 3894 GGTCTGGGTGAGAGCTACTCCGAGAGGGCGGTGCTGTGGCGGCGGCTCTTCCGGGGC 3953
QY 935 acaggtgagcgagcgagcggtgacacgcgacgagcgagcgagcgagcgagcgagcg 991
DB 3954 AGCGGGGCTGCTGCGAGAGTGGCTCTTCCGAGAGCAGCGCGGTGCTGCGAGGA 4010

RESULT 7
US-08-458-240-1
Sequence 1, Application US/08458240
Patent No. 6143727

GENERAL INFORMATION:

APPLICANT: Roop, Dennis R.
APPLICANT: Rothnagel, Joseph A.
APPLICANT: Greenhalgh, David A.
TITLE OF INVENTION: SPECIFIC EXPRESSION VECTORS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90017

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458.240
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/146,930
FILING DATE:
APPLICATION NUMBER: 07/876,286
FILING DATE: April 30, 1992
APPLICATION NUMBER: No. 6143727 yet assigned (204/132)
FILING DATE: October 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Waidburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 204/152
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6530 bases

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-458-240-1

Query Match 2.9%; Score 49.8; DB 3; Length 6530;
Best Local Similarity 50.6%; Pred. No. 0.0022;
Matches 120; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 755 gcccttcacagacacagcgagcggtcgtgacacgcgtcgcgcgacgagcgagcgagcggt 814
DB 3774 GCCCCCCAGAGAGCTACGAGAGGGGCTCTCCGAGAGAGGTGACTCTGAGAGTGGC 3833
QY 815 acacgcgagatcgacaggtgagcgagcggtgacacgcgacgacgagcgagcgagcg 874
DB 3834 TCCTCTGGCGGCGGTGGCGGCGGTGCTCTACTCCAGCGGTGTGGGCGGAGCAGCGGT 3893
QY 875 ggtgatactccgacgcgacagcggtgagcgagcggtgagcgagcggtgagcgagcg 934
DB 3894 GGTCTGGGTGAGAGCTACTCCGAGAGGGCGGTGCTGTGGCGGCGGCTCTTCCGGGGC 3953
QY 935 acaggtgagcgagcgagcggtgacacgcgacgagcgagcgagcgagcgagcgagcg 991
DB 3954 AGCGGGGCTGCTGCGAGAGTGGCTCTTCCGAGAGCAGCGCGGTGCTGCGAGGA 4010

RESULT 8
PCT-US93-03993-1
Sequence 1, Application PC/TUS9303993

GENERAL INFORMATION:

APPLICANT: Roop, Dennis R.
APPLICANT: Rothnagel, Joseph A.
APPLICANT: Greenhalgh, David A.
TITLE OF INVENTION: CONSTITUTIVE AND INDUCIBLE EPIDERMAL
TITLE OF INVENTION: VECTOR SYSTEMS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77010-3095

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03993
FILING DATE: 19930428
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Paul, Thomas D.
REGISTRATION NUMBER: 32,714
REFERENCE/DOCKET NUMBER: D-5405
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713/651-5325
TELEFAX: 713/651-5246
TELEX: 762829

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 6530 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US93-03993-1

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2001, 20:28:34 ; Search time 1342.69 Seconds
(without alignments)
12172.567 Million cell updates/sec

Title: US-09-597-513-1
Perfect score: 1729
Sequence: 1 tccactgcgtcatttgaa.....tgaataaagggttgacctc 1729

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
```

```
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 556)
AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Broksstein, P., Lewis, S. and Rubin, G.M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
Plate: 421 row: H column: 7
High quality sequence stop: 525.
Location/Qualifiers
1..556
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD42191"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XLI Blue"
/note="Organ: embryo; Vector: pOT2; Site_1: EcorI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2."

BASE COUNT 174 a 206 c 104 g 72 t
ORIGIN

Query Match
Best Local Similarity 3.3%; Score 56.8; DB 21; Length 556;
Matches 118; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy 773 ggcggcctcggtacacccgtcgccgcatagcgggcgccggtacacccggtatgcagcagt 832
Db 229 ggtggttcacgttcgtggtggtccggtgagagaggtggttcaacttgcgtggtccggt 170
Qy 833 ggcggcgggtgtatagcgcacagcgagcggtggtggtggtatctcgcagcgca 892
Db 169 ggagaggtggttcaactttagtggctggtggtggtggtggtggtggtggtggtggt 110
Qy 893 acagcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 952
Db 109 ggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 50
Qy 953 ggcacacccactgcaacagcggtggtggtggtggtggtggtggtggtggtggtggt 992
Db 49 ggcttaattgt 10

RESULT 3
CNS04HMV/c 630 bp DNA GSS 21-MAY-2000
LOCUS Tetradon nigroviridis genome survey sequence PUC-ori end of clone
DEFINITION 110P12 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL291136.1 GI:8029716
VERSION AL291136.1
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 630)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Fitzames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 630)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fitzames, C., Winkler, P., Brotlier, P., Quetier, F., Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 630)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetradon>.
Location/Qualifiers
1..630
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="110P12"
/clone_1id="G"
/note="Genoscope sequence ID : COBGL10DC06SP1-end : PUC-ori"

BASE COUNT 121 a 198 c 113 g 158 t 40 others
ORIGIN

Query Match
Best Local Similarity 3.2%; Score 55.2; DB 221; Length 630;
Matches 129; Conservative 14; Mismatches 146; Indels 0; Gaps 0;

Qy 704 accactccataataaagcagagacacccatagagacagctgtatgcagctcccttc 763
Db 573 AACACACACAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 514
Qy 764 cagacacacggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggtcggt 823
Db 513 AAANNATAANNNNNCAANANAGCGGAGGTTGCGGCGCGGCGGCGGCGGCGGCGGCGG 454
Qy 824 ggcagaggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 883
Db 453 RAARKMGTGKTKTGCGGWRGCGGTGAGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 394
Qy 884 ccgacgcacacggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 943
Db 393 GCGCGTGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 334
Qy 944 ggcagcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 992
Db 333 GGTGCGGTGGTGAGAGGTGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 285

RESULT 4
CNS0091P/c 925 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACRI9D16 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013.1 GI:4934461
VERSION AL053013.1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

```
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
```


Query Match	2.98;	Score 50;	DB 221;	Length 975;
Best Local Similarity	55.18;	Pred. No. 0.013;	80;	Indels 0;
Matches 98;	Conservative	0;	Mismatches	0;

db 145 GGGGCGCCGCCCATGTCTAAGGCGCATTGCACCGCGCCGCCGCCATTCCTTGAAGCCCCCA 20

926 cccactgcacacagttgcygcagcgtg 954
|
| | | | | | | | | | | | | | | |


```
/dev_staging-elimixyo
```

UI-R-AI and UI-R-EI libraries. The UI-R-AI library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta adult lung

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2001, 21:42:44 ; Search time 44.08 Seconds
(without alignments) 583.135 Million cell updates/sec

Title: US-09-597-513-2

Perfect score: 2200
Sequence: 1 MSIGTRPPQRTTPLDPSA.....MTDVKHAKDKTQASTQHTL 424

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: A_Geneseq.0601.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2200	100.0	424	20	AAW97851
2	2200	100.0	424	21	AAW71098
3	2200	100.0	424	21	AAW84859
4	559	25.4	447	20	AAW96260
5	559	25.4	447	21	AAW71094
6	559	25.4	447	21	AAW84855
7	318	14.5	197	19	AAW83014
8	318	14.5	197	19	AAW7412
9	294.5	13.4	221	20	AAW28446
10	215	9.8	730	21	AAW29582
11	215	9.8	752	21	AAW29581

12	201	9.1	402	8	AAW70709	Plasmodium cynomol
13	179.5	8.2	941	19	AAW68205	M. catarrhalis str
14	177.5	8.1	2870	21	AAW95559	Caenorhabditis ele
15	177.5	8.1	3178	21	AAW95556	Caenorhabditis ele
16	176	8.0	447	21	AAW29728	Arabidopsis thalia
17	176	8.0	468	21	AAW29727	Arabidopsis thalia
18	171.5	7.8	564	22	AAW70182	Peptide dendrimer
19	171.5	7.8	892	19	AAW68203	M. catarrhalis str
20	166.5	7.6	287	11	AAW05877	Merzite surface a
21	163.5	7.4	344	18	AAW06600	Hypersensitive res
22	163.5	7.4	344	19	AAW75865	Pseudomonas solana
23	163.5	7.4	344	19	AAW62457	Pseudomonas solana
24	163.5	7.4	344	19	AAW61116	Hypersensitive res
25	163.5	7.4	344	20	AAW67641	A hypersensitive r
26	163.5	7.4	344	21	AAW71099	Pseudomonas solana
27	163.5	7.4	344	21	AAW4860	A hypersensitive r
28	162.5	7.4	898	18	AAW1853	Mycobacterium tube
29	161.5	7.3	440	21	AAW9354	Human PRO1411 (UNG
30	161.5	7.3	440	22	AAW87551	Human PRO1411. HO
31	161.5	7.3	440	22	AAW81205	Amino acid sequenc
32	161.5	7.3	440	22	AAW6103	Protein of the inv
33	161	7.3	258	21	AAW51723	Arabidopsis thalia
34	160.5	7.3	831	19	AAW68201	M. catarrhalis str
35	159.5	7.2	263	21	AAW36620	Arabidopsis thalia
36	159.5	7.2	273	21	AAW36619	Arabidopsis thalia
37	159.5	7.2	309	21	AAW36618	Arabidopsis thalia
38	159	7.2	388	20	AAW04999	Mycobacterium spec
39	159	7.2	873	19	AAW68207	M. catarrhalis str
40	158	7.2	647	18	AAW0926	Human cyclin D1-hu
41	158	7.2	647	19	AAW4575	Human cyclin D1-hu
42	158	7.2	705	18	AAW18573	Human cyclin D1-hu
43	158	7.2	705	18	AAW14574	Human cyclin D1-hu
44	156.5	7.1	749	17	AAW4895	Human trophoblast
45	156.5	7.1	749	21	AAW33357	Human trophoblast

ALIGNMENTS

RESULT 1	
AAW97851	
ID	AAW97851 standard; Protein: 424 AA.
AC	AAW97851;
DT	07-JUN-1999 (first entry)
XX	
DE	Hypersensitive response elicitor HrpW.
XX	
KW	Hypersensitive response elicitor; dspe gene; HrpW;
XX	transgenic plant; disease resistance; insect resistance.
OS	Pseudomonas syringae pv. tomato.
XX	
FT	Key
FT	Domain
FT	Location/Qualifiers
FT	1..186
FT	/note="hypersensitive response elicitor-like
FT	domain"
FT	Region
FT	119..186
FT	/note="region of 6 imperfect glycine-rich repeats"
FT	187..424
FT	/note="C-terminal domain"
PN	W09907207-A1.
XX	
PD	18-FEB-1999.
XX	
PF	24-JUL-1998; 98WO-US15501.
XX	
PR	06-AUG-1997; 97US-0055107.
XX	
PA	(CORR) CORNELL RES POUND INC.
XX	

PI Alfano JR, Charkowski A, Collmer A;
XX WPI: 1999-167125/14.
DR N-PSDB; AAX24347.
XX
XX New hypersensitive response eliciting (dspe) gene and protein
PT useful for providing transgenic plants and seeds with enhanced
PT growth, and insect and disease resistance
XX
XX Claim 18; Page 38-39; 56pp; English.
XX
XX This the amino acid sequence of hypersensitive response element
CC (HRE) HrpW of *Pseudomonas syringae* pv. tomato DC3000, as deduced
CC from the nucleotide sequence of an isolated DNA molecule (see
CC AAX24347). HrpW includes an HRE-like domain that is rich in Gln,
CC Ser and Gly and which includes 6 imperfect glycine-rich repeats
CC with many acidic and polar residues that align with similar
CC repeats in the HrpZ proteins of *P. syringae* pv. *syringae* and *P.*
CC *syringae* pv. *tomato*. The amino acid sequence of this region
CC suggests alternating beta-sheets and turns that may form a barrel
CC structure. The C-terminal domain of HrpW is similar to several
CC fungal and bacterial Pst proteins. The HrpW protein or isolated
CC DNA molecule can be used to impart disease resistance to plants, to
CC enhance plant growth and/or to control insects on plants. This is
CC achieved by applying the HRE protein in a non-infectious form to
CC plants or plant seeds. Alternatively, transgenic plants or plant
CC seeds transformed with DNA encoding the HRE can be provided.
XX
XX Sequence 424 AA:
SQ
Query Match 100.0%; Score 2200; DB 20; Length 424;
Best Local Similarity 100.0%; Pred. No. 7.5e-155;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSIGITRPQQTTPPLDPSALSGSKSPQNTFGEONTQQAIDPSALLFGSDTKDVFNGTP 60
DB 1 msigilrppqqtllpdlfalslsgskspntfgeqntqgaidsallfgsdtkdvnifgtp 60
QY 61 DSTVONFODASKPNDSOSNTAKLISALIMSLQMLTNSNKKOPDNOPOSOAPFQNNGG 120
DB 61 dstvnpqpdaskpndsgsnlaklislalmsllqmltnsnkkqdtngqpsqpfqnnng 120
QY 121 LGTPSADSGGGGTPDATGGGGGDTSPATGGGGGDTPTATGGGGGGGTPATGGSGSGT 180
DB 121 lgtpsadsggggtpdatgggggdtspatgggggdtptatgggggggtpatggsgsgt 180
QY 181 PRTAGGEGGVTPQITPOLANPRTSGTGSVSDPTAGSTBOAGKINVKDTIKVAGAEVFD 240
DB 181 prtagggegvtptitpolanpntsgtgsvsdptagstboagkinnvkdtkvgaevfd 240
QY 241 GHGATFTADSKSMNGDGENOKPMFELAEAGATLKNNVLGENEVDGIHVAKNAQEVITDN 300
DB 241 ghgattfadsksmngdgenokpmfelaegatlknvnlgenevdgihvaknaevitdn 300
QY 301 VHAQNWGEDLITVKGEGGAATVNLNTKNSAKAGADKVVOLNANTHLKIDNFRADFGTM 360
DB 301 vhaqnvgedliltvkggegaaavtnlnknsakagadkvvvqlnanthlkidnfradfgtm 360
QY 361 VRTNGKQOPDMSIELNGTEANHGKFAIVKSDSDDLKLTAGNTAMDVKAAYDKTOASTQ 420
DB 361 vrtngkqpfidmsielngteanhgkfaivksdsddlklatgnlamtdvkhaydktoastq 420
QY 421 HTEEL 424
DB 421 htel 424

XX 08-SEP-2000 (first entry)
DT Pseudomonas syringae dspe gene encoded hypersensitive response elicitor.
XX
XX DE Hypersensitive response elicitor; environmental stress resistance;
XX KM plant; pathogen; dspe gene.
XX KM
XX OS *Pseudomonas syringae*.
XX WO200028055-A2.
XX PN
XX PD 18-MAY-2000.
XX
XX PF 04-NOV-1999; 99NO-US26039.
XX PR 05-NOV-1998; 98US-0107243.
XX PA (EDEN-) EDEN BIOSCIENCE CORP.
XX
XX PI Wei Z, Schading RL;
XX WPI: 2000-376566/32.
XX DR N-PSDB; AAD00673.
XX
XX PT Application of a hypersensitive response elicitor protein to plants to
PT impart stress resistance
XX
XX PS Disclosure; Page 25-26; 84pp; English.
XX
XX The patent discloses a method to impart stress resistance to plants by
CC applying a hypersensitive response elicitor in a non-infectious form to
CC a plant or seed. The present sequence is a hypersensitive response
CC elicitor protein encoded by dspe gene from *Pseudomonas syringae*.
CC The protein elicits a plant pathogen's hypersensitive response and is
CC used to impart stress resistance to plants.
XX
XX Sequence 424 AA:
SQ
Query Match 100.0%; Score 2200; DB 21; Length 424;
Best Local Similarity 100.0%; Pred. No. 7.5e-155;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSIGITRPQQTTPPLDPSALSGSKSPQNTFGEONTQQAIDPSALLFGSDTKDVFNGTP 60
DB 1 msigilrppqqtllpdlfalslsgskspntfgeqntqgaidsallfgsdtkdvnifgtp 60
QY 61 DSTVONFODASKPNDSOSNTAKLISALIMSLQMLTNSNKKOPDNOPOSOAPFQNNGG 120
DB 61 dstvnpqpdaskpndsgsnlaklislalmsllqmltnsnkkqdtngqpsqpfqnnng 120
QY 121 LGTPSADSGGGGTPDATGGGGGDTSPATGGGGGDTPTATGGGGGGGTPATGGSGSGT 180
DB 121 lgtpsadsggggtpdatgggggdtspatgggggdtptatgggggggtpatggsgsgt 180
QY 181 PRTAGGEGGVTPQITPOLANPRTSGTGSVSDPTAGSTBOAGKINVKDTIKVAGAEVFD 240
DB 181 prtagggegvtptitpolanpntsgtgsvsdptagstboagkinnvkdtkvgaevfd 240
QY 241 GHGATFTADSKSMNGDGENOKPMFELAEAGATLKNNVLGENEVDGIHVAKNAQEVITDN 300
DB 241 ghgattfadsksmngdgenokpmfelaegatlknvnlgenevdgihvaknaevitdn 300
QY 301 VHAQNWGEDLITVKGEGGAATVNLNTKNSAKAGADKVVOLNANTHLKIDNFRADFGTM 360
DB 301 vhaqnvgedliltvkggegaaavtnlnknsakagadkvvvqlnanthlkidnfradfgtm 360
QY 361 VRTNGKQOPDMSIELNGTEANHGKFAIVKSDSDDLKLTAGNTAMDVKAAYDKTOASTQ 420
DB 361 vrtngkqpfidmsielngteanhgkfaivksdsddlklatgnlamtdvkhaydktoastq 420
QY 421 HTEEL 424

|||||
Db 421 htel 424

RESULT 3
AAV84859
ID AAV84859 standard; Protein: 424 AA.

XX AAV84859;

XX 08-AUG-2000 (first entry)

XX A hypersensitive response elicitor protein.

XX Hypersensitive response; insect control; disease resistance;

XX Hypersensitive response elicitor; plant growth; vegetable; crop;

XX ornamental plant.

XX Pseudomonas syringae.

XX WO200020452-A2.

XX 13-APR-2000.

XX 05-OCT-1999; 99WO-0523181.

XX 05-OCT-1998; 98US-0103050.

XX (EDEN-) EDEN BIOSCIENCE CORP.

XX Wei Z, Fan H, Niggemeyer JI;

XX WPI: 2000-303745/26.

XX N-PSDB: AAA14943.

XX Hypersensitive response elicitor polypeptides useful for imparting

XX enhanced growth, disease resistance and insect resistance to plants,

XX especially vegetables and ornamental flowers.

XX Disclosure: Page 26-28; 100pp; English.

XX The present sequence represents a hypersensitive response elicitor

XX polypeptide. The specification describes hypersensitive response

XX elicitor polypeptide fragments, which do not elicit a hypersensitive

XX response. Instead, the proteins impart disease resistance to plants,

XX enhance plant growth, and/or control insects. The polypeptide

XX fragments may be used to these properties to plants. The plants which

XX may be treated in this way include vegetables, crops and ornamental

XX plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,

XX peanut, corn, potato, sweet potato, bean, pea, chitoy, lettuce,

XX endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,

XX broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,

XX carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,

XX citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,

XX tomato, sorghum or sugarcane. Arabidopsis thaliana, Saintpaulia,

XX petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.

XX Sequence 424 AA;

Query Match 100.0%; Score 2200; DB 21; Length 424;

Best Local Similarity 100.0%; Pred. No. 7.5e-155;

Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSIGITPPQQTTPPLDFSAISGKSPONTGEONTQOAIIDPSALIFGSDTQKDVNFGTP 60

Db 1 msigltppqgttltpldtsalsgskspnttgeqntqaidpsallfisdqkdvntfpt 60

QY 61 DSTVONPDASKPNDOSNAIKLISALIMSLQMTNSNKKQDTONQOEPDSDOAFPNNGG 120

Db 61 dstvnpdqaskpndsgnialklsalimsllqmtlnsnkkqdtneqdpdsqafpnngg 120

QY 121 LGTPSADSGGGTPTATGGGGGDFPSATGGGGGDFPTATGGGSGGGTPTATGGSGGT 180

|||||
Db 121 lgtpeadsqgggttptatgsggggtdtpeatsqgggtdtptatgsgsggttptatgsgsgt 180

QY 181 PTATGGGEGVTPQITPOLANPNRTSGTSVSDTAGSTEQAKINVKOTIKVAGAEVFD 240

Db 181 ptatgggegvtptqitpolanpnrtsgtsvdsdtagsteqaklnvkdtklvagaevfd 240

QY 241 GHGATFTADKSMKNGDQENOKPMFELAEATLKRVNIGENPVDSIHKKANAOEVTIDN 300

Db 241 ghgattadkmsmgngdqqenqkpmfelaeatlkrvnlgenevdghkknagaevtldn 300

QY 301 VHAONVGEDLITVKEGGAATVNLIRKNSAKGADKVVQIANHILKIDNFKADDFCTM 360

Db 301 vhaonvgedliltvkeggaaatvnlirknssakgaddkvvqianhilkidnfdkddftm 360

QY 361 VRTNGKQFDDMSIELNCEANHGKFAVLKSDSDDLKATGNIAMTVKHAVDKTOASTQ 420

Db 361 vrtngkqfddmsielnglneanhgkfalvkdsddlklatgniamtdvkhaqdktgastq 420

QY 421 HTEL 424

Db 421 htel 424

RESULT 4

AAW96260

ID AAW96260 standard; Protein: 447 AA.

XX AAW96260;

XX 14-JUN-1999 (first entry)

XX Hypersensitive response eliciting protein (HrpH).

XX Hypersensitive response elicitor protein; hypersensitive response;

XX hrpW; pathogen; infection; crop protection; disease resistance;

XX pest resistance; transgenic plant; colouration; maturation.

XX Erwinia amylovora.

XX WO9907208-A1.

XX 18-FEB-1999.

XX 27-JUL-1998; 98WO-US15547.

XX 06-AUG-1997; 97US-0055108.

XX (CORR) CORNELL RES FOUND INC.

XX Beer SV, Kim JF;

XX WPI: 1999-167126/14.

XX N-PSDB: AAX09007.

XX New Erwinia amylovora hypersensitive response eliciting gene and

XX protein - useful for providing transgenic plants and seeds with

XX enhanced growth, and insect and disease resistance

XX Claim 1; Page 50-51; 54pp; English.

XX The hypersensitive response eliciting protein (hrp) or polypeptide

XX is produced as part of an active defense by plants against

XX incompatible pathogen infections. The hypersensitive response is a

XX rapid localised necrosis. The hrp protein and gene when used in

XX nucleotide constructs are useful for providing disease resistance to

XX plants, insect control to plants, and enhancing plant growth

XX (enhancing fruit size and earlier colouration and maturation), by

XX direct application of the protein to plants, or by producing

XX transgenic plants or seeds using the hrp gene.

XX Sequence 447 AA;

CC response elicitor protein from *Erwinia amylovora*. The protein is

PT especially vegetables and ornamental flowers -

PF 09-APR-1998; 98EP-0106586.
 XX
 PR 08-SEP-1997; 97JP-0242735.
 PR 09-APR-1997; 97JP-0091142.
 XX
 PA (KAOS) KAO CORP.
 XX
 PI Hatada Y, Ito S, Kobayashi T, Koike K, Suzumatsu A;
 PI Yoshimatsu T;
 XX
 DR WPI, 1998-523159/45.
 DR N-PSDB; AAV59478.
 XX
 PT New Bacillus pectic acid lyase - useful as a detergent component, a
 PT food-processing agent and a fibre-processing agent
 XX
 PS Claim 1, Page 16-17; 29pp; English.
 XX
 CC The present sequence represents a pectic acid lyase isolated from
 CC microorganism Bacillus sp. KSM-P15. The pectic acid lyase has high
 CC pectic acid lyase activity which degrades pectin in plant cell walls
 CC and fibre in vegetables, and so is useful as a component of detergents,
 CC a food-processing agent, or a fibre-processing agent. The pectic acid
 CC lyase has a higher optimum reaction pH (10.3-10.7) than known Bacillus
 CC pectic acid lyases (pH 8-9.5) and so has wider industrial applications.
 CC Unlike present pectic acid lyases, the new enzyme has a high enzyme
 CC activity, and can be produced on a mass scale.
 XX
 SQ Sequence 197 AA;

Query Match 14.5%; Score 318; DB 19; Length 197;
 Best Local Similarity 40.7%; Pred. No. 3,5e-16;
 Matches 74; Conservative 33; Mismatches 61; Indels 14; Gaps 5;
 QY 226 VAKDTIVGAGEVFDGATFTAD-KSMGNGDGENCKPMELEAGATLNKVNIGENEVD 284
 DB 4 VVHETIVPGQGTGDTGQTYVAVNPGLIGDGSQENQKPIFLIAGASLKNVIVGAPAD 63
 QY 285 GIVKAKNAQAEVTDNVAQNVGEDLTIVGEGGAATNINIKKSSAKGADKVVQANAN 344
 DB 64 GVNHCY----DCLTVYVWEDVGEDALTLKSSG-----TVLISGSAAYKAYDKVFQINAA 114
 QY 345 THKIDNFKADDEFTWRTGKQFDMSEIANGIEANHGKFAVKSDDSLKATGNTA 404
 DB 115 GTLNIHFRFDDIGKIVTQNGTLY-KVMMVENCISRYKDALIRDS----STSTGRIV 170
 QY 405 MT 406
 DB 171 nt 172

RESULT 9
 AAT28446
 ID AAY28446 standard; Protein; 221 AA.
 AC AAY28446;
 XX
 DT 17-JAN-2000 (first entry)
 DE Bacillus licheniformis Pectate lyase I.
 XX
 KW Pectate lyase I; EC 4.2.2.2; pectin degrading enzyme; ATCC 14580;
 KW catalytic active domain; cellulose binding domain; CBD; operably linked;
 KW optimum activity; pH; detergent composition; yarn; cellulosic fibre;
 KW recycled waste paper; pulp; retting process; animal feed; wine; juice;
 KW transgenic plant.
 XX
 OS Bacillus licheniformis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..27
 FT /label= Pro-sequence

FT Protein 28..221
 FT /label= Mature_pectate_lyase_I
 FT MISC-difference 133
 FT /note= "Conserved residue"
 FT MISC-difference 155
 FT /note= "Conserved residue"
 XX
 PN WO9927083-A1.
 XX
 PD 03-JUN-1999.
 XX
 PF * 24-NOV-1998; 98WO-DK00514.
 XX
 PR 24-NOV-1997; 97DK-0001344.
 PR 06-MAY-1998; 98US-0073684.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Andersen LN, Bjornvad ME, Lange NEK, Schnorr K, Schuelein M;
 XX
 DR WPI, 1999-610578/52.
 DR N-PSDB; AAX89484.
 XX
 PT New isolated pectate lyase enzymes -
 XX
 PS Claim 6; Page 79-80; 93pp; English.
 XX
 CC The present sequence is pectate lyase I, which is a pectin degrading
 CC enzyme derived from Bacillus licheniformis, ATCC 14580. The enzyme
 CC comprises a catalytic active domain and a cellulose binding domain
 CC (CBD) that are operably linked to each other. It shows optimum activity
 CC at pH greater than 9 and temperature 55 degree centigrade. It can be used
 CC in detergent compositions, for cleaning hard surfaces, for machine
 CC treatment of fabrics, for improving the properties of cellulosic fibres,
 CC yarn, woven or non-woven fabric, for the degradation of plant material
 CC e.g. recycled waste paper, mechanical paper-making pulps or fibres
 CC subjected to retting process, for preparing animal feed and for
 CC processing wine or juice. DNA encoding the enzyme can also be used for
 CC the production of transgenic plants.
 XX
 SQ Sequence 221 AA;

Query Match 13.4%; Score 294.5; DB 20; Length 221;
 Best Local Similarity 34.6%; Pred. No. 2,2e-14;
 Matches 72; Conservative 38; Mismatches 71; Indels 27; Gaps 5;
 QY 219 EQAGKINVVKDTIKVAGEVFDGATFTADKSMGNGDGENCKPMELEAGATLNKVN 278
 DB 24 EKALAAEVVHKLIVVEKGYDGGKRLIAGPELIGDSQFREDGKPIKVEDGATLKNV 83
 QY 279 GENEVDGIVHAKNAQAEVTDNVAQNVGEDLTIVGEGGAATNINIKKSSAKGADK 338
 DB 84 GAPADGVHLYG---NASINNVWEDVGEDALTVSSEG-----SVLISGSAARLADK 134
 QY 339 VOLNANTHLKIDNFKADDEFTWRTGKQFDMSEIANGIEANHGKFAVKSDDSLK 398
 DB 135 FGINKASTFTVKHFTADGGKFLIQLGYSFTKAV-VNIDCITLTKMEALIRDS----- 188
 QY 399 ATGNIAATD-----VKHAYDK 414
 DB 189 STSVTNTLTRYSKVGGKWIGVKHATER 216
 RESULT 10
 AAG29582
 ID AAG29582 standard; Protein; 730 AA.
 AC AAG29582;
 XX
 DT 17-OCT-2000 (first entry)
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 35222.

XX Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX Arabidopsis thaliana.
OS EPI033405-A2.
PN
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144326.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145214.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

```
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 9.8%; Score 215; DB 21; Length 730;

Best Local Similarity 24.0%; Pred. No. 8e-08;

Matches 99; Conservative 46; Mismatches 180; Indels 88; Gaps 17;

```
QY 25 SPQNTGEBNTQOALIDPSALLFGSDTQKDVNFGT---PSTYONPDA--SKPNDQSN 79
Db 127 apppgstgqak-----pgasgigsdsigsagtlpqaqtrcteknagskpsssag 179
QY 80 IAKLISALIMSLQMLTNKSKKODTNEQDPSQAPFQNNGLGT-PSADSGGGGTPATG 138
Db 180 tnpgaasa-----vngsteknagskps-----gsaflnpasasgungeteknv 224
QY 139 GG-----GGDPNSATGGGGGDTPTATGGGGGGG--TPATGGGGGTPATGGGE 188
Db 225 gsgkpsqgkgtlpgnagngtlnkagskassgsaartlmpasaggetvsnlgtde 284
QY 189 ---GGTTPQITPOLANPNKISGTSVSDTAGSTEQ---ACKINVKDT-LKVGAGEVFDG 241
Db 285 snaggy-----sksnpganngasglsnagstglfagagtgigtgleesdagsgsktng 337
QY 242 HGATFPAADKSGMGNDGGENCKPMFELAEGATLKNVNLGENEVDGIIHKAKNAQEVITDNY 301
Db 338 nggtlndgaasigsndsgtctnpg---aggtcdsnlsgtlenngyket-----np 383
QY 302 HAONVGEDLITVKGEGGAATNLIKNSAKGADKVVOLNANTHLKIDNFKADDFGTWV 361
```

```
Db 384 gasigig-----nsdsgstgspegtesnadgtkltntgkesntgsesntnspqkleaq 436
QY 362 RTNGGKOFDDMSIELNGICENHGKFPALVKSDSDDLKLTATNTMTDVKKAYDK 414
Db 437 gngngnqwd-d-gtchdgyvmkih-----vaygigleqtrfdyvk 474
```

RESULT 11

AA629581 ID AA629581 standard; Protein; 752 AA.

AA629581;

17-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 35221.

Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

23-APR-1999; 99US-0130891.

28-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

30-APR-1999; 99US-0132407.

04-MAY-1999; 99US-0132484.

05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.

07-MAY-1999; 99US-0132487.

11-MAY-1999; 99US-0132863.

14-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

18-MAY-1999; 99US-0134370.

18-MAY-1999; 99US-0134768.

19-MAY-1999; 99US-0134941.

20-MAY-1999; 99US-0135124.

21-MAY-1999; 99US-0135353.

24-MAY-1999; 99US-0135629.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.

28-MAY-1999; 99US-0136782.

01-JUN-1999; 99US-0137222.

03-JUN-1999; 99US-0137502.

07-JUN-1999; 99US-0137724.

08-JUN-1999; 99US-0138094.

10-JUN-1999; 99US-0138540.

10-JUN-1999; 99US-0138847.

14-JUN-1999; 99US-0139119.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2001, 21:21:44 ; Search time 26.43 Seconds
(without alignments)
330.318 Million cell updates/sec

Title: US-09-597-513-2

Perfect score: 2200
Sequence: 1 MSIGITPPRQOTTPLDPSA.....MTDVKHAIDKTOASTOHTEL 424

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCBUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2200	100.0	424	US-09-120-817-2
2	318	14.5	197	US-09-402-668-2
3	294.5	13.4	221	US-09-198-856-4
4	169	7.7	201	US-09-032-995-1
5	169	7.7	201	US-09-053-003-40
6	169	7.7	269	US-08-452-531-4
7	169	7.7	269	US-08-460-746A-4
8	169	7.7	269	US-08-460-555-4
9	169	7.7	269	US-08-460-066-4
10	163.5	7.4	344	US-08-891-234-7
11	163.5	7.4	344	US-08-819-539-7
12	163.5	7.4	344	US-09-030-270A-7
13	163.5	7.4	344	US-08-984-207-7
14	163.5	7.4	344	PCT-US96-08819-7
15	158	7.2	647	US-08-770-761A-8
16	158	7.2	705	US-08-770-761A-7
17	156.5	7.1	749	US-08-317-522A-2
18	156.5	7.1	749	US-08-439-818A-2
19	156.5	7.1	749	US-08-751-965-2
20	156.5	7.1	749	US-08-738-975-2
21	156.5	7.1	749	US-08-728-626-2
22	156.5	7.1	749	US-08-808-599A-2
23	153.5	7.0	235	US-08-529-190B-1
24	153.5	7.0	318	US-09-060-756-727
25	152.5	6.9	482	US-08-358-160-5
26	152.5	6.9	484	US-08-358-160-7
27	152.5	6.9	907	US-09-010-928B-4

28	151.5	6.9	62	4	US-09-402-668-10	Sequence 10, Appl
29	150.5	6.8	674	1	US-08-317-522A-3	Sequence 3, Appl
30	150.5	6.8	674	1	US-08-439-818A-3	Sequence 3, Appl
31	150.5	6.8	674	1	US-08-751-965-3	Sequence 3, Appl
32	150.5	6.8	674	2	US-08-738-975-3	Sequence 3, Appl
33	150.5	6.8	674	2	US-08-728-626-3	Sequence 3, Appl
34	150.5	6.8	674	3	US-08-808-599A-3	Sequence 3, Appl
35	150	6.8	872	2	US-08-387-942C-5	Sequence 5, Appl
36	149.5	6.8	420	2	US-08-845-998-8	Sequence 8, Appl
37	149.5	6.8	420	4	US-09-206-537-8	Sequence 8, Appl
38	149	6.8	461	2	US-08-463-587A-26	Sequence 26, Appl
39	149	6.8	461	2	US-08-463-667A-4	Sequence 4, Appl
40	149	6.8	461	5	PCT-US91-09133-27	Sequence 25, Appl
41	149	6.8	461	5	US-09-051-09133-27	Sequence 27, Appl
42	148	6.7	334	4	US-09-060-756-728	Sequence 728, App
43	147	6.7	2763	3	US-08-496-944-2	Sequence 2, Appl
44	146.5	6.7	1160	3	US-08-808-599A-24	Sequence 24, Appl
45	146	6.6	211	1	US-08-276-852-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-09-120-817-2
Sequence 2, Application US/09120817
Patent No. 6172184
GENERAL INFORMATION:
APPLICANT: Colimer, Alan
APPLICANT: Charkowski, Amy
APPLICANT: Alfano, James R.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
TITLE OF INVENTION: PSEUDOMONAS SYRINGAE AND ITS USE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/120, 817
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/055,107
FILING DATE: 06-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1741
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 424 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-09-120-817-2
Query Match 100.0%; Score 2200; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 1,7e-172;
Matches 424; Conservative 0; Mismatches 0; Indels 0;

```

QY 1 MSIGITPRPOQTTPPLDPSALSGKSPQNTFGEONTOAIDPSALLFGSPTOKOVNFGTP 60
DB 1 MSIGITPRPOQTTPPLDPSALSGKSPQNTFGEONTOAIDPSALLFGSPTOKOVNFGTP 60
QY 61 DSTVONPDASKPNDSQSNIAKLISALIMSLQMLTNSNRKODTNOBQPSQAPFQNNNG 120
DB 61 DSTVONPDASKPNDSQSNIAKLISALIMSLQMLTNSNRKODTNOBQPSQAPFQNNNG 120
QY 121 LGTPSADSGGGGTPDANGGCGGTPSATGGGGGTPTATGGGGGGGTPTATGGGGGT 180
DB 121 LGTPSADSGGGGTPDANGGCGGTPSATGGGGGTPTATGGGGGGGTPTATGGGGGT 180
QY 181 PTATGGGEGVTPQITPQLANPNRTSGSVSDTASSTEOAGKINVKQDTIKVAGAEVFD 240
DB 181 PTATGGGEGVTPQITPQLANPNRTSGSVSDTASSTEOAGKINVKQDTIKVAGAEVFD 240
QY 241 GHGATFTADKSMKNGDQGENQKPMFEIABGATLKNNVLGENEVDGIHVAKNAQEVITDN 300
DB 241 GHGATFTADKSMKNGDQGENQKPMFEIABGATLKNNVLGENEVDGIHVAKNAQEVITDN 300
QY 301 VHAONVGEDLITVKGEGGAVTNLTNNSAKGADKVVQNLNMTHLKIDNFKRADDGCTM 360
DB 301 VHAONVGEDLITVKGEGGAVTNLTNNSAKGADKVVQNLNMTHLKIDNFKRADDGCTM 360
QY 361 VRTNGKQFDDMSIELNGIEANHGKFPALVKSDDDLKATGNTAMTDVKHAYDKTQASTQ 420
DB 361 VRTNGKQFDDMSIELNGIEANHGKFPALVKSDDDLKATGNTAMTDVKHAYDKTQASTQ 420
QY 421 HTL 424
DB 421 HTL 424

```

```

RESULT 2
US-09-402-668-2
; Sequence 2, Application US/09402668
; Patent No. 6172030
; GENERAL INFORMATION:
; APPLICANT: WADA, Yasunao
; APPLICANT: KASAI, Miyuki
; APPLICANT: SHIKATA, Shitsuw
; APPLICANT: SUZUMATSU, Atsushi
; APPLICANT: KOIKE, Kenzo
; APPLICANT: HATADA, Yuji
; APPLICANT: KOBAYASHI, Tohru
; APPLICANT: ITO, Susumu
; APPLICANT: TSUMADORI, Masaki
; TITLE OF INVENTION: Detergent Composition
; FILE REFERENCE: 2173-0116P
; CURRENT APPLICATION NUMBER: US/09/402,668
; CURRENT FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 9-091142 JAPAN
; PRIOR FILING DATE: 1997-04-09
; PRIOR APPLICATION NUMBER: 9-242736 JAPAN
; PRIOR FILING DATE: 1997-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/01613
; PRIOR FILING DATE: 1998-04-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Bacillus sp.
; OTHER INFORMATION: Strain: KSM-P15
US-09-402-668-2

```

```

Query Match 14.5%; Score 318; DB 4; Length 197;
Best Local Similarity 40.7%; Pred. No. 7e-19;
Matches 74; Conservative 33; Mismatches 61; Indels 14; Gaps 5;
QY 226 VVVDITKVGAGEVFDHGATFTAD-KSMGNGDQGENQKPMFEIABGATLKNNVLGENEVD 284

```

```

DB 4 VVHETITVPAGQGFEDGKGQYVYANPNLTGDSQAENQKPIFLREAGASLKNVYIGAPAD 63
QY 285 GIHVAKNAQEVITDNVHAONVGEDLITVKGEGGAVTNLTNNSAKGADKVVQNLN 344
DB 64 GHVGYG---DCTITNTWEDVEDALTLKSSG---TVNIGGAAYKRYDVFQINAA 114
QY 345 THLIKIDFKADDFGTWRTNGKQFDDMSIELNGIEANHGKFPALVKSDDDLKATGNTA 404
DB 115 GTINIRFRADIDGLKYKONGGITY-KVYANVENCNISRYKDALIRDS---STIGRIY 170
QY 405 MT 406
DB 171 NT 172

```

```

RESULT 3
US-09-198-956-4
; Sequence 4, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schmorl, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-198-956-4

```

```

Query Match 13.4%; Score 294.5; DB 4; Length 221;
Best Local Similarity 34.6%; Pred. No. 6.9e-17;
Matches 72; Conservative 38; Mismatches 71; Indels 27; Gaps 5;
QY 219 EQAGKINNVKDTIKVAGEVFDHGATFTADKSMGNGDQGENQKPMFEIABGATLKNNVL 278
DB 24 EKALAEVHVAKTIVKGGYDGGKRLIAGPELGDQSGREDQKPIFKVBDGATLKNNVL 83
QY 279 GENEVDGIHVAKNAQEVITDNVHAONVGEDLITVKGEGGAVTNLTNNSAKGADK 338
DB 84 GAPADGVHTYG---NASINNVMEDVGDALTVSEG---SVTINGSARLADKI 134
QY 339 VOLANTHLKIDFKADDFGTWRTNGKQFDDMSIELNGIEANHGKFPALVKSDDDLK 398
DB 135 FQINKASTFTVKNFPTADQSGKFTROLGSGFTFKAV-VNINDCTTNKKEALFRIDS----- 188
QY 399 ATGNIAMTD-----VXHA YDK 414
DB 189 STSSVTMTNTRYSKVGOKWIGVKAHATER 216

```

```

RESULT 4
US-09-052-995-1
; Sequence 1, Application US/09052995
; Patent No. 6183956
; GENERAL INFORMATION:
; APPLICANT: SivaraJa, Mohanram
; APPLICANT: Strulovici, Berta
; APPLICANT: Flores, Osvaldo A.

```



```

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-448 MIS:bh
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-452-531-4

Query Match      7.7%; Score 169; DB 1; Length 269;
Best Local Similarity 31.9%; Pred. No. 1.7e-06;
Matches 58; Conservative 14; Mismatches 68; Indels 42; Gaps 10.

QY 109 PDSOAPFQNNG-----GLTPNSAD----SGGGT--PDATGGGGGDPPSATGGGGGDT 155
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 89 PGSECPKGTGPGSBPGKGTGPGSECGKGTGPGSEPKGTGTGGSESPKGTGTGGSGEG 148
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 156 PRATGGGGS-----GGGGT--PRATGGSGGTPTATGG-----GEGGVNPIITPOLA 200
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 149 PKGTGGPSEBPKGTGGGSGEKGPGSGESPKGTGGPGESEPKGTGGPGESE -GPKGT 207
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 201 NPNTSTGGSVDPAAGSIEQAKIIVVMDTIKVGADEVFDGCAIFTLADKSMGNNDQGN 260
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 208 GKKTGGPSEBKEGEP-----KGT--GGPGESEBKEGPKGTGGPGSG-GESHSH 254
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 261 QK 262
   | :
Db 255 KK 256

RESULT 7
US-08-460-746A-4
; Sequence 4, Application US/08460746A
; Patent No. 5780591
GENERAL INFORMATION:
APPLICANT: MATLASHEMSKI, Gregory
APPLICANT: CHAREST, Hugues
TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED LEISHMANIA
TITLE OF INVENTION: GENES AND PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,746A
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,463
FILING DATE: 12-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,987
FILING DATE: 03-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973

```

```

: REFERENCE/DOCKET NUMBER: 1038-483 MIS:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 269 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-460-746A-4

Query Match          7.7%; score 169; DB 1; Length 269;
Best Local Similarity 31.9%; Pred. No. 1.7e-06;
Matches 58; Conservative 14; Mismatches 66; Indels 42; Gaps 10;

QY 109 PDSQAPFQNNG-----GLTPPSAD---SGGGGT--PDATGGGGGDPSPATGGGGDT 155
      10 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
DB 89 PSESEPKGTGCGSPGSPKGTGGPGSEGPKGTGPGSEGPKGTGCGSPGSPKGTGCGSPG 148
      156 PRTATGGGGG-----GGGGT--PRTATGGGGGTPRTATG-----GGGGYPTITPOLA 200
      149 PRTGTGPGSEGPKGTGCGPGSPGPKGTGGPGSPGPKGTGCGPGSEGPKGTGCGPSE--GPKGT 207
QY 201 NENRTSGTGSVDATGSTEDQAGKINVVADTIKVGAGEYFDGHATFTADKSMGNDQGEN 260
      208 GPKGTGPGPSEAGTEGCP-----KGT--GPGSEAGTEGPKGTGPGSG--GEHSHN 254
QY 261 QK 262
      255 KK 256

RESULT
      8
US-08-460-555-4
: Sequence 4, Application US/08460555
: Patent No. 5827671
: Patent No. 5827671 5827671
: GENERAL INFORMATION:
: APPLICANT: MATLASHWSKI, Gregory
: APPLICANT: CHAREST, Hugues
: TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED LEISHMANIA
: TITLE OF INVENTION: GENES AND PROTEINS
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Slim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/460,555
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,463
: FILING DATE: 12-SEP-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/115,987
: FILING DATE: 03-SEP-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-485 MIS:vg

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-460-555-4

Query Match 7.7%; Score 169; DB 2; Length 269;
Best Local Similarity 31.9%; Pred. No. 1.7e-06;
Matches 58; Conservative 14; Mismatches 68; Indels 42; Gaps 10;

QY 109 PDSQAPFQNNG-----GLGTPSAD---SGGGT--PDATGGGGGDTPPSATYGGGGDT 155
DB 89 PGSEBPKGTGGPGSGPKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEB 148
QY 156 PTATGGGGS-----GGGGT--PTATGGGSGGTPTATGG-----GGGVTPTQITPOLA 200
DB 149 PKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEBPKGT 207
QY 201 NPNRTSGTGSYDPAAGSTEQAKINVKDTIKVAGEYFDGATFTADKSMGNDGGEN 260
DB 208 GPKGTGGPGSEAGTEGP-----KGT--GGPGSEAGTEGPRTGTGPGSG-GEHSHN 254
QY 261 OK 262
DB 255 KK 256

RESULT 9
US-08-460-066-4
Sequence 4, Application US/08460066
Patent No. 6133017
GENERAL INFORMATION:
APPLICANT: MATIASHEMSKI, Gregory
APPLICANT: CHAREST, Hugues
TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED LEISHMANIA
TITLE OF INVENTION: GENES AND PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & Mcburney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,066
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,463
FILING DATE: 12-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,987
FILING DATE: 03-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-484 MIS:V9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-460-066-4

Query Match 7.7%; Score 169; DB 4; Length 269;
Best Local Similarity 31.9%; Pred. No. 1.7e-06;
Matches 58; Conservative 14; Mismatches 68; Indels 42; Gaps 10;

QY 109 PDSQAPFQNNG-----GLGTPSAD---SGGGT--PDATGGGGGDTPPSATYGGGGDT 155
DB 89 PGSEBPKGTGGPGSGPKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEB 148
QY 156 PTATGGGGS-----GGGGT--PTATGGGSGGTPTATGG-----GGGVTPTQITPOLA 200
DB 149 PKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEBPKGT 207
QY 201 NPNRTSGTGSYDPAAGSTEQAKINVKDTIKVAGEYFDGATFTADKSMGNDGGEN 260
DB 208 GPKGTGGPGSEAGTEGP-----KGT--GGPGSEAGTEGPRTGTGPGSG-GEHSHN 254
QY 261 OK 262
DB 255 KK 256

RESULT 10
US-08-891-254-7
Sequence 7, Application US/08891254
Patent No. 5776889
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: Hypersensitive Response
TITLE OF INVENTION: Induced Resistance in Plants
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/891,254
FILING DATE: 10-JUL-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/475,775
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 14603/10050
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-891-254-7

Query Match 7.4%; Score 163.5; DB 1; Length 344;
Best Local Similarity 23.2%; Pred. No. 6.6e-06;
Matches 88; Conservative 31; Mismatches 153; Indels 107; Gaps 15;

QY 1 MSIGITRPPQQTTPIDFSLSGSKSPQPNTEFGQNTQA-----IDP 42
DB 1 MSVG-----NIOSPSNLPGLQNLINLTNTNSOQSGSVODLIKQVEKDLITLITIALVOK 54
QY 43 SALFGSDT-----QKD--VNFETPD-----STVONPDASKP--NDSQSNIAKLISA 86
DB 55 AASAGAGTNTGNTGNAAPKADONANAGANDPSPKNSQAPQASANKTGNVDANNQDPMQA 114
QY 87 LMSLLQMLTNSKKQDPTNOEPDSPAPOFONGGLGTPSADSGGGGTPDRTG----- 138
DB 115 L-MQLLEDLVKL-LKALHMQOPGNDKNGVGGANGANGAGGQGLAEALQETIEQLAQ 172
QY 139 -GGGGDTPSATGGGGGDTPTATGGGSGGGGTPATGGGSGGTPATGGEGGVTPTP 197
DB 173 LGGGAGAGAGAGAGAGAGAGADGSGAGAGANGADGNGVNGNQANG-----P 222
QY 198 QLANPNTSGTGSVSDTASGTEQAGKINVKYDTIKY----- 233
DB 223 Q--NAGDVNANGADD--GSEDDGGLTGLVQLKMLKLITLALVQMMQGGGLGGNQAGGSK 278
QY 234 GAGEVDPGHATFTADKSMKNGDQGENQKMFELAEAGATLKNNLGENEVDGIHYAKNA 293
DB 279 GAGNASPASGANPGANGPGSADDOSSQN-----NLSQIMDVV---KEY 320
QY 294 QEVTIDNVHAQNVGEDLIT 312
DB 321 VQILQMLAQAQNGSSQST 339

RESULT 11

US-08-819-539-7
Sequence 7, Application US/08819539
Patent No. 5859324

GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: Hypersensitive Response
TITLE OF INVENTION: Induced Resistance In Plants
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.

ZIP: 14603

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/819,539

FILING DATE: 17-MAR-1997

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/475,775

ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.

REGISTRATION NUMBER: 30,727

REFERENCE/DOCKET NUMBER: 14603/10050

TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-819-539-7

Query Match 7.4%; Score 163.5; DB 2; Length 344;
Best Local Similarity 23.2%; Pred. No. 6.6e-06;
Matches 88; Conservative 31; Mismatches 153; Indels 107; Gaps 15;

QY 1 MSIGITRPPQQTTPIDFSLSGSKSPQPNTEFGQNTQA-----IDP 42
DB 1 MSVG-----NIOSPSNLPGLQNLINLTNTNSOQSGSVODLIKQVEKDLITLITIALVOK 54
QY 43 SALFGSDT-----QKD--VNFETPD-----STVONPDASKP--NDSQSNIAKLISA 86
DB 55 AASAGAGTNTGNTGNAAPKADONANAGANDPSPKNSQAPQASANKTGNVDANNQDPMQA 114
QY 87 LMSLLQMLTNSKKQDPTNOEPDSPAPOFONGGLGTPSADSGGGGTPDRTG----- 138
DB 115 L-MQLLEDLVKL-LKALHMQOPGNDKNGVGGANGANGAGGQGLAEALQETIEQLAQ 172
QY 139 -GGGGDTPSATGGGGGDTPTATGGGSGGGGTPATGGGSGGTPATGGEGGVTPTP 197
DB 173 LGGGAGAGAGAGAGAGAGADGSGAGAGANGADGNGVNGNQANG-----P 222
QY 198 QLANPNTSGTGSVSDTASGTEQAGKINVKYDTIKY----- 233
DB 223 Q--NAGDVNANGADD--GSEDDGGLTGLVQLKMLKLITLALVQMMQGGGLGGNQAGGSK 278
QY 234 GAGEVDPGHATFTADKSMKNGDQGENQKMFELAEAGATLKNNLGENEVDGIHYAKNA 293
DB 279 GAGNASPASGANPGANGPGSADDOSSQN-----NLSQIMDVV---KEY 320
QY 294 QEVTIDNVHAQNVGEDLIT 312
DB 321 VQILQMLAQAQNGSSQST 339

RESULT 12

US-09-030-270A-7
Sequence 7, Application US/09030270A
Patent No. 5977060

GENERAL INFORMATION:
APPLICANT: Zilber, Thomas A.

APPLICANT: Wei, Zhong-Min

TITLE OF INVENTION: INSECT CONTROL WITH A

TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.

ZIP: 14603

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,270A

FILING DATE: 28-FEB-1997

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/039,226
ATTORNEY/AGENT INFORMATION:


```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/08819
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/475,775
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/10051
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-08819-7

```

```

Query Match          7.4%; Score 163.5; DB 5; Length 344;
Best Local Similarity 23.2%; Pred. No. 6.6e-06;
Matches 88; Conservative 31; Mismatches 153; Indels 107; Gaps 15;

```

```

QY 1 MSIGTIPRPOQTTPIDFSLSGSKSPQNPFPFGQNTQQA-----IDP 42
DB 1 MSG-----NIOPSRLPGLQNLNLTNTNSQSGSVODLLKQYEKDLNIIALYOK 54
QY 43 SALLFSDT-----OKD--VNFETPD-----STVQNPQASKP-NDQSNIKLISA 86
DB 55 AASAGAGNGTNGTNAKADKDNANAGANDPSKNDPSKSOAPQANKTGVNDANNQPMQA 114
QY 87 LMSLQMLTNSKKQDNTNEOPDPSQAPFQNNGLGTPSADSGGCTPDATG----- 138
DB 115 L-MQLLEDVLKL-LKALAHMQPGKNDKNGVGGANGANGAGGGLAELEIEQILAQ 172
QY 139 -GGGADTPSATGGGGDTPATGSGSGGGGCTPTATGSGSGGTPTATGGEGVTPQITP 197
DB 173 LGGGAGAGAGAGGAGGAGADGSGAGGAGANGADGNGVNGANG-----P 222
QY 198 QLANPNTSGTGSVDTAGSTEQAGKINVKTKIV----- 233
DB 223 Q--NAGDVANGANGAD--GSEPDGGLTGLVQLKILMLALVQMGGGLGGGNAQGGSK 278
QY 234 GAGEVDGATGATTAKSMNGDQGENOKMFELAGATLKNVNGENVDGIHVAKKA 293
DB 279 GAANASPASANGANGAPGADQSSGQN-----NLSQQLMDVY---KEY 320
QY 294 QEVYIDNVHAQNVGEDLIT 312
DB 321 VOILQMLAANGSGQST 339

```

```

RESULT 15
US-08-770-761A-8
; Sequence 8, Application US/08770761A
; Patent No. 5814503
; GENERAL INFORMATION:
; APPLICANT: Kovacevic, Steven
; APPLICANT: Otto, Keith A.
; APPLICANT: Rao, Ramachandra N.
; TITLE OF INVENTION: FUSION PROTEINS COMPRISING CELL CYCLE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company

```

```

; STREET: Lilly Corporate Center/Patent Division
; City: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,761A
; FILING DATE: 19-DEC-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10136
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-376-0756
; TELEFAX: 317-277-1917
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-770-761A-8

```

```

Query Match          7.2%; Score 158; DB 2; Length 647;
Best Local Similarity 24.6%; Pred. No. 4.4e-05;
Matches 99; Conservative 49; Mismatches 140; Indels 114; Gaps 23;

```

```

QY 14 TPLDF-SALSGSPQNPTEGEQNTQQAIDPSALLFESDTQKQVNF-----GTPDS 62
DB 156 TPHDFIEHFLSKME---AEN-KQIRKHAQTFALCATDVKFTSNPMSVAAGSVVA 210
QY 63 TVQNPQDASKPND--SQSNIAKLISALI-----MELQMLTNSNKK 101
DB 211 AVQG-LNLRSPNNFLSYRLTRFLSHVIRKCDPDCLRACQEQIEALLESLRQAQNMMDK 269
QY 102 QDTPNOBQPSQAPFQNNGLGTPS-----ADSGGCTPDATGGGGDTPSATGGGGD 154
DB 270 AAEFEFEFEFE---EVDLACTPTDVRVDIASMGG-----SGGSG-----GSSG- 314
QY 155 TPTATGGGSGGGGCTPTATGSGSGGTPTATGGGGGVTPQITPOLANPNTSGTGSVDT 214
DB 315 ---GSGGSGGSGGSLSGKGGGSG-----GGGSG----- 341
QY 215 AGSTEQAGKINVKTKIKYA-GEVF---DGAGATTADKS--MGNGDQGENOKPMFELA 268
DB 342 -GGSMATSRPEYAE-IGVAGTGYVKARDPHSGHFVAKSVAPVGGGGGLPISTVR 399
QY 269 EGATLKNVNGEN---EVDGIHVAKNAQEVYIDNVHAQNVGEDLITYKGE---GGA 320
DB 400 EVALLRLAFEPHPNVYRLMDVYATSRTRDKITLVF-EHVDQDLRTYLDKAPPGGLPA 458
QY 321 VTNLNTRKSSAKGADKVVQLNANTH--LKIDNFKADFGTM 360
DB 459 EFKIDLMRQFLRGLD--FLHANCIVHRDLKPERILVTSGTV 498

```

Search completed: October 4, 2001, 22:26:02
Job time: 3858 sec

Fri Oct 5 10:04:25 2001

us-09-597-513-2.ra1

Page 9

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2001, 22:23:59 ; Search time 40.64 Seconds

(without alignments)
794.734 Million cell updates/sec

Title: US-09-597-513-2

Perfect score: 2200

Sequence: 1 MSIGITPRPQQTTPRLDPSA.....MTDVKHAIDKQASQTHREL 424

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

PIR 68: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	559	25.4	447	2	T18447
2	291.5	13.2	221	2	A70045
3	216	9.8	3190	2	T13828
4	215	9.8	730	2	F96559
5	205	9.3	219	2	S68364
6	204.5	9.3	680	2	T08080
7	198.5	9.0	1901	2	F70806
8	197.5	9.0	401	1	OZQAC
9	196.5	8.9	1226	2	T24045
10	194	8.8	3016	2	S77300
11	193	8.8	2038	2	A43742
12	191.5	8.7	526	1	S33799
13	189.5	8.6	1489	2	D70807
14	189	8.6	528	2	G02127
15	188	8.5	667	2	A70893
16	187.5	8.5	813	2	S70795
17	186.5	8.5	622	2	I37984
18	185.5	8.4	1079	2	B70807
19	185	8.4	207	2	T07381
20	183	8.3	731	2	C70974
21	182.5	8.3	1538	2	H70846
22	181.5	8.2	1381	2	E70806
23	179.5	8.2	694	2	F70868
24	178.5	8.1	242	2	A45724
25	178.5	8.1	741	2	G70917
26	177.5	8.1	388	2	JC6164
27	177.5	8.1	491	2	D70916
28	177.5	8.1	1113	2	S28925
29	177.5	8.1	2329	2	T28125

30	176.5	8.0	1585	2	T31611	hypothetical prote
31	176.5	8.0	1844	2	T51890	related to Nup98-N
32	176	8.0	447	2	G84687	probable disease r
33	176	8.0	618	2	A70989	hypothetical glycl
34	176	8.0	1306	2	A70934	hypothetical glycl
35	175	8.0	572	2	T08509	trbl protein - Ent
36	175	8.0	853	2	A70896	hypothetical glycl
37	174.5	7.9	570	2	S07330	keratin, epidermal
38	174.5	7.9	2232	2	T34434	hypothetical prote
39	173.5	7.9	615	2	H70589	hypothetical glycl
40	173.5	7.9	995	2	T22942	hypothetical prote
41	173	7.9	302	2	C84470	hypothetical glycl
42	173	7.9	2059	2	D82671	surface protein XF
43	172.5	7.8	481	2	A35628	loricrin - mouse
44	172.5	7.8	801	2	F70824	hypothetical glycl
45	172.5	7.8	1329	2	E70917	hypothetical glycl

ALIGNMENTS

RESULT 1
T18447
HrpW protein - Erwinia amylovora
C:Species: Erwinia amylovora
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:Gaudriault, S.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z18936
A:Accession: T18447
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-447 <GAU>
A:Cross-references: EMBL:Y13831; NID:e1293450; PID:e1293451; PIDN:CA74158.1
C:Experimental sources: strain CFBP1430; specific host Pommoidae
C:Genetics:
A:Note: hrpW

Query Match	25.4%	Score 559;	DB 2;	Length 447;
Best Local Similarity	36.1%	Pred. No. 1.7e-24;		
Matches 147;	Conservative 54;	Mismatches 132;	Indels 74;	Gaps 13;
QY 72	KPNDQS--NIAKLISLMSLOMTNSNKKODTNOEQDPDSQAPFONGSLG-----	122		
DB 37	QPIDRQTEQMAQLALRLKSL--LSPOSNAATGAGNDQTTGVNAGGLNGRKGTAG 93			
QY 123	-TPSADS-----GGGG-----TPDATGGGG-GDTP-----SATGGG 151			
DB 94	TTPOSDSQNMISEMNGNGLDQATTPDQGGQIDNPLRLAKMLKLRAMDGSDQDFGP 153			
QY 152	GGDPPTATGGGSGGCGTPTATGGG---SGGTPATGGGEGVTPQTPOL-----A 200			
DB 154	GTGNSSASGSSSSGSGPNDLSGKAPSGNSPCGNSPVSTFSPSPSPSPSPSPSPS 213			
QY 201	NPNTSG-----TGSVSDTASG---TEQAKINVKTIYVAGCEVVDGAT 245			
DB 214	SPTKAGSGTPVDPHPDVPVSGAGISAGANSVAFTSAGANQVTLHDTITVKAQGVDDGQOT 273			
QY 246	FTAKSMGNGDGENOKPMFELAGATLLKNVNLGENEVDGIHVAKNAQETIDINVAON 305			
DB 274	FTASELIDGGQSENQKPLLEDGASLKNVWGDDGADGHLTG---DKIDNLVATN 329			
QY 306	VEGDLITVKGEGGAAYVNLINKNSAKGADKVVOLNANHTLKIDNFKADDFGTVRTNG 365			
DB 330	VEGALITVKNKSNACKRSHVETNSSFHASDKILQNLADNPLSDVNKAKDFGTVRTNG 389			
QY 366	GKQFDMSIELNGIEANHGKPAIVKSDSDIKLATGNIAMTVKHA 412			
DB 390	GQD-GNMDLNLHSIADRGKFSFVKSDSEGLNVTSDISLDVENHY 435			

RESULT 2

A70045

pectate lyase homolog yvpa - *Bacillus subtilis*

C:Species: *Bacillus subtilis*

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999

C:Accession: A70045

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.: Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fedret, C.; Ferrati, E.
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Gallazzi, A.; Gall
leach, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.
Koeffer, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: lauber, J.; Iazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schroefer, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033

A:Accession: A70045

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-221 <KUN>

A:Cross-references: GB:299121; GB:AL009126; NID:92635827; PIDN:CA15500.1; PID:ell186183;

A:Experimental source: strain 168

C:Genetics:

A:Gene: yvpa

Query Match 13.2%; Score 291.5; DB 2; Length 221;

Best Local Similarity 36.6%; Pred. No. 7.3e-10;

Matches 68; Conservative 32; Mismatches 71; Indels 15; Gaps 4;

QY 226 VVKDITKVGAGFEDHSGATFTADKSMGNGDGENQKPMFELEGATILKNVNGENVEDG 285

DB 31 VHERITIVKNTTYDCKGRFVAGKELGDSQSENDPFRVVDGATILKNVYVGAARAAG 90

QY 286 HVKAKNAOEVTIDYNAONVGEDLITVKGEGAAVTNINIKNSAKGADKRVYOLNANT 345

DB 91 VHTYG---NVNIQNKWEDVGEDALTVEKGE---KVTIDGSAOKASDKIFQINKAS 141

QY 346 HIKINERKADDCGTWRTGSGKQFDMSEINIEANHKFKALVKSDDSLKATFENIM 405

DB 142 TTTVKNFTADNGKFTRLQJGSGTF-HVDVITDKCTTNMKEALFRTDS---KTSTVR 195

QY 406 TDVKHA 411

DB 196 TMTYRS 201

RESULT 3

T13828

CREB-binding protein homolog - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000

C:Accession: T13828

R:Akamaru, H.; Chen, Y.; Dai, P.; Hou, D.X.; Nonaka, M.; Smolik, S.M.; Armstrong, S.; GC
Nature 386, 735-738, 1997

A:Title: *Drosophila* CBP is a co-activator of cubitus interruptus in hedgehog signalling.

A:Reference number: 217785; MUID:97263578

A:Accession: T13828

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-3190 <AKT>

A:Cross-references: EMBL:U88570; NID:91916929; PID:91916930; PIDN:AA53050.1

C:Genetics:

A:Cross-references: FlyBase:FBgn0015624

A:Map position: X

C:Superfamily: bromodomain homology

F:173-1780/Domain: bromodomain homology <BRO>

Query Match 9.8%; Score 216; DB 2; Length 3190;

Best Local Similarity 25.1%; Pred. No. 0.00028;

Matches 105; Conservative 37; Mismatches 147; Indels 130; Gaps 21;

QY 14 TPLDFSLGSKSPQPTFEON--TQOALDPSALLGSPDQKDVNCFDPTVQNPQDAS 71

DB 1305 TPLSLEALNAGAPGPTGGSASNVYSAFSPGFLSNP-----STGTP-----SNNNS 1356

QY 72 KPNDOSNTAKLISALIMSLQMLTNSKKODTNOBPDSPQAFQNNGL--GTP----- 124

DB 1357 SAANNPPSVSLMQ-----DPLSN---RPGPPYIPASVPATASGLAASVTPASAA 1406

QY 125 ----SADSGGGTPDATTGGGGGDTPSATGGGG-----DTPTA-----TGGGGGGGGT 169

DB 1407 ATCCASSGSSNSSGATAGASSTSSSSAGSGTPLSVSTPSATMATSSGGGGGGG 1466

QY 170 PVATGGSGGTPTA-----TGCGGGVTPQIT----- 196

DB 1467 ---AGGSSSTTPASNLMLMSGTACGGTGATTTTSSSSKRMSSSSLSQMALLEA 1523

QY 197 -----POLANPNRTSGTSVSDTAGSTEQACKINVK--DTIKVAGEVFDGHA 244

DB 1524 ARNDDETPSPSGENTNGSGSGN---AGMASKGLKIDSIKDDDIK-----K 1568

QY 245 TPTADKSMGNGDGENQKPMFELEGATILKNV-----LGENVYDGIHVAKNA 293

DB 1569 EFMDDSCGGNDSO---MDCSTGGGKKNVNDGTSIMKEIKETEDIGD-EVYIK-T 1622

QY 294 QEVTIDYNAONVGEDLITVKGEGAAV---TNINIKNSAKGADKRVYOLNANTHL 348

DB 1623 EAMDVDEAGGSTAGEH---HGEGGGSGVGGKKNINGAHOGAIGAVNDIKPKETIK 1677

RESULT 4

F96559

hypothetical protein F5F19.6 [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: F96559

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Matzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; MUID:21016719

A:Accession: F96559

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-730 <STO>

A:Cross-references: GB:AE005173; NID:94220464; PIDN:AA012691.1; GSPDB:GN00141

C:Genetics:

A:Gene: F5F19.6

A:Map position: 1

Query Match 9.8%; Score 215; DB 2; Length 730;

Best Local Similarity 24.0%; Pred. No. 5.8e-05;

Matches 99; Conservative 46; Mismatches 180; Indels 88; Gaps 17;

QY 25 SPQNTFGFQNTQOALDPSALLGSPDQKDVNCFDPTVQNPQDAS--SKPNDOSN 79

DB 127 APPGSTGAK-----PASGIGSDSGSISGAGTNPADGTFETKKNAGSKPSSGSA 179

QY 80 IAKLISALIMSLQMLTNSKKODTNOBPDSPQAFQNNGLGCT-PSADSGGGGCTPDATG 138

DB 180 TNPGASA-----VNGETETERNAGSKRPS-----GSAGTNPGASAGNGETEKV 224

RESULT 6
T08080
probable myrosinase-binding protein - rape
N:Alternate names: jasmonate inducible protein
C:Species: Brassica napus (rape)
C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 11-May-2000
C:Accession: T08080
R:Geshi, N.; Brandt, A.
Plantla 204, 295-304, 1998

Query Match 9.08; Score 198.5; DB 2; Length 1901;

```

Best Local Similarity 29.2%; Pred. No. 0.0015;
Matches 86; Conservative 13; Mismatches 87; Indels 109; Gaps 14;

OY 119 GGLGTPSADSGGGTPDATGGGGDTPSATYGGGG---DTPATG-----GGSGGG 167
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 711 GGEAGAGCNSGVGTGSSGAGGAGCGCTGGAGGSGADNPACGAGGAGCGAAGAG 770

OY 168 GMPRTATG-GGSGCTPTATG-----GGEGGVTPOTPOLANPNRRTSGTG----- 209
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 771 GAGGATGTGTGTGGVVGATGTSAGIGGAGGRGSDG-----DGAISGLISGPD 818

OY 210 ---SYSDPAGSTEDQAKINVVVDTIKVGAGEVFDDGATFTADKSMGNSDGENOKPMFE 266
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 819 GGGGGGGAGGAGSAGACGIN-----GAG-----GAG-----GNGGDG----- 849

OY 267 LAEGATLKNNVIGENEVDGIHVAKANNAOEVTIDNVAQNVEEDLITVKEGGAAYTNLNI 326
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 850 -GDGAT-GAAGLGIDNGVGVDGAGAGAG-----NGNAGVGLTAKADGGAAGNGNG 901

OY 327 KNSSAKGADDKVQVVLNANTHLKIDNFKADDFG-----TWVRTNGG 366
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 902 GAGGAGGAGD-----NNFNGGGAGGAGGGGGGLGAGASTISINANGG 943

RESULT 8
OZQZNC
circumsporozoit protein precursor - Plasmodium cynomolgi (strain Gombak)
N:Alternate names: major sporozoite surface antigen
C:Species: Plasmodium cynomolgi
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 07-Nov-1997
C:Accession: E26255
R:Galinski, M.R.; Annot, D.E.; Cochran, A.H.; Barnwell, J.W.; Nussenzweig, R.S.; Enea,
Cell 48, 311-319, 1987
A:Title: The circumsporozoite gene of the Plasmodium cynomolgi complex.
A:Reference number: A90889; MUID:87102878
A:Accession: E26255
A:Molecule type: DNA
A:Residues: 1-401 <GAL>
C:Comment: There are three distinct regions in the mature circumsporozoite protein, the
obscure membrane-anchoring sequence.
C:Comment: There are 17 tandem copies of the 11-residue repeat D/G-A-A-A-G-G-G-G-N.
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C:Keywords: sporozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-401/Product: circumsporozoite protein #status predicted <MAT>
F:98-278/Region: 11-residue repeats
F:326-379/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 9 0%; Score 197.5; DB 1; Length 401;
Best Local Similarity 33.5%; Pred. No. 0.00028;
Matches 60; Conservative 12; Mismatches 76; Indels 31; Gaps 6;

OY 119 GGLGTPSADSGGGTPD---ATGGGGDTPSATYGGGGTPATYGGG-----SSGGGP 169
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 170 GGGGMDGAAAGAGGGGNDGAAAGAGGGGNDGAAAGAGGGGNDGAAAGAGGGG 229

OY 170 PTATGGSGGCT--PTATGGGEGGVTPOTPOLANPNRRTSGTSVSDPAGSTEDQAKINVV 227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230 DGAAGAAAGGGGNDGAAAGAGGNGGAA--AGGGGNGGAAAGGGGNGGAAAGG 280

OY 228 KDTIKVGAAGEVFDGHAFTTADKSMGNSDGENOKPMFELAECAITLKNNVIGENEVDGI 286
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 281 -----RAGDQDPPAGNKKAGEAGGNAGAGGQGN-----NEGANVPNPKIVLEYLDKI 328

RESULT 9
T24045
hypothetical protein R08B4.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T24045
C:White, S.

```

submitted to the EMBL Data Library, November 1995

A:Reference number: Z19834

A:Accession: 124045

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1226 <MBL>

A:Cross-references: EMBL:Z68008; PIDN:CAA92000.1; GSPDB:GN00028; CESP:R08B4.1

A:Experimental source: clone R08B4

C:Genetics:

A:Gene: CESP:R08B4.1

A:Map position: X

A:Introns: 53/2; 113/1; 152/3; 204/3; 272/1; 354/1; 389/3; 714/3; 839/3; 877/1; 960/3

Query Match 8.8%; Score 196.5; DB 2; Length 1226;

Best Local Similarity 26.9%; Pred. No. 0.0012;

Matches 94; Conservative 24; Mismatches 156; Indels 75; Gaps 14;

QY 3 IGTPRRP-----QQTPPLDPSALSG-----KSPDNPFGE 33

Db 710 IGRTPYPHONVYNDIGSAKPPNCPADISLGSMAACDGGIQLFEANKYIDAKOPEPFDK 769

QY 34 QNTQQAIDPSALFEGSDIQKDVNFETPDPSTVQNPQDAS-KPNDSQSNIAKLIALMSLL 92

Db 770 HNTRL-----AKIMRAVQK-RREGTFESVVALEADFEWGNKFKFRCKIDS----- 816

QY 93 QMLTNSNKKQDTNOEQPDSQ-APFQNNGLCTPSADSGGG-----TPDATGGGGGDTFSA 147

Db 817 -QGVNALTIVYQSSSKPPSPSDIDIPNDPTLGGPGSSGGGGGGGSGGGGGGGGGSG 875

QY 148 TGGGGGDTPLATGGGGSGGGGTPPATGGSGGTPPATGGGGGGVTPQTLPLDANPRTSG 207

Db 876 GSGGGGSSNSGGGGGGNGGGG-----NGGGNGNGGGAGADGNGG-----AGAGNGNG 922

QY 208 TGSVSDTGSTGTEOAKINVVYDITKVGAEVFDHGATFTADKSKNGCD-OGENOKPWF 266

Db 923 AG-----AGNGNGAGAGN-----GNGAGAG-----NGNGAGAGNGNGAGADASAAAAAQ 970

QY 267 LAEGGTLKNVLGENEVDGIHYKAKNAOEYITIDNVHAQNVEDLITVKG 315

Db 971 AAAAAQAQAAAAAAQAQAAAAAAQAQAAAAAAQAQAAAAAANPLSALVAATG 1019

RESULT 10

S77300

hypothetical protein slr1403 - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S77300

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

A:Reference number: S74322; MUID:97061201

A:Accession: S77300

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-3016 <KAN>

A:Cross-references: EMBL:D90907; GB:AB001339; NID:91652618; PIDN:BA17634.1; PID:d101

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 8.8%; Score 194; DB 2; Length 3016;

Best Local Similarity 20.9%; Pred. No. 0.0046;

Matches 118; Conservative 63; Mismatches 180; Indels 204; Gaps 25;

QY 22 SGKSPQPTFGEQNTQAIDPSALIFGS-----DTQKDVN-----FGTPDSTVQ 65

Db 2408 AGAANSPT-----IDMTGIAAGTIILNSPLPTLNNDITFTVTLGAFTPTKFN 2458

QY 66 NPQASRKNDSQSNIAKLIALMSLLQMLTNSNKKQDTNOEQPDSQA----- 113

```

Db 2459 NP-----SAFNIIETLEGEVAVNLSTYMNLLONGYAKGGDGVSGGGGAGAGALTIIQGS 2511
QY 114 -----PFRONN-----GGIGTPSADSGG--GTPD--ATGGGG-----142
Db 2512 TVIIDNVSFSDNRAVGGNGTAGARGGGEIFGTPDDRPSAGGGGNFNASGNANNGAAG 2571
QY 143 -----DTPBATTGGGGDPTATAGGGGSGG-----GTP--TATGGSGGTPTATG-----185
Db 2572 SVGDKAKAGCTGCTGTGGTGGGGGGGGSSGSPDFTPRNGSGGPGAGGFGAGG 2631
QY 186 -----GEGGVTP-----193
Db 2632 GAGGGGGSSGQNPGRDCKRGFGSGSGGNPIFGNGSVSGGAGTVEGGSGNGGS 2691
QY 194 -----QITPOLANPMTSGTSVSDPTAGEA--GKINVKDTIKVAGE 237
Db 2692 GGGAGLGLGALFINQAVTITN-SQFSGNTTQGGTGNSGALGALFITDNSFVTAGL 2750
QY 238 VFDDHGATFTADKSMNG-DGQENQKPMFELAEATLKNVL-----GENVDCI- 286
Db 2751 TFSNNSAPSSPGSGFSNAYGGFQNNNDYGTIDLSNASTMLFLVPLLPPEVQHEISVYA 2810
QY 287 -----HYAKNAOEVTIDNVHQAQNVGEDLTIVKGGG-----AAVTNLINIKSSAKGAD- 335
Db 2811 KDNLFVQLPDGSSEVSL-----LYEDQPFVSGSGFQMWQILLEAETINGVANNVLMQNPDI 2863
QY 336 DKVQVLANTH--LKIDNFKADDEGTM-----VRTNGKQKQPDMSIELNGIEANHGK 385
Db 2864 DEIVGVNADSNMNMNISEETPTNSFTLEAEIFQIDLNGELLCD---RLTNIISK-GN 2919
QY 386 FALVKSDDSLDKLATGNAMTDVHK 410
Db 2920 TSFLEGIPTGNTYVQTGDLTAPIKY 2944

RESULT 11
A43742
female sterile homeotic protein, 205k - fruit fly (Drosophila melanogaster)
N:Alternate names: membrane protein fsh, 205k
N:Contains: female sterile homeotic protein, 110k
C:Species: Drosophila melanogaster
C>Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 20-Sep-1999
R:Haynes, S.R.; Mozer, B.A.; Bhatia-Dey, N.; David, I.B.
Dev. Biol. 134, 246-257, 1989
A:Title: The Drosophila fsh locus, a maternal effect homeotic gene, encodes apparent men
A:Reference number: A43742; MUID:89276730
A:Accession: A43742
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2038 <RAY>
A:Cross-references: EMBL:M23221; NID:9157452; PIDN:AAA28540.1; PID:9157453
A:Accession: B43742
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1106 <MA2>
A:Cross-references: EMBL:M23222
C:Genetics:
A:Gene: fsh
A:Cross-references: FlyBase:FBgn0004656
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
C:Keywords: alternative splicing; transmembrane protein
F:1-2038/Product: female sterile homeotic protein, 205k #status predicted <MA2>
F:1-1106/Product: female sterile homeotic protein, 110k #status predicted <MA2>
F:59-116/Domain: bromodomain homology <BRO1>
F:503-560/Domain: bromodomain homology <BRO2>

```

Query Match 8.8%; Score 193; DB 2; Length 2038;
 Best Local Similarity 24.0%; Pred. No. 0.0033;
 Matches 80; Conservative 40; Mismatches 133; Indels 80; Gaps 14;

```

QY 36 TQOAIIDPSALLFGSDPKDVNFGP-----DSTVONPODASKPNDSQNIAKLISALIMS 90
Db 1721 TSAVAQAATAATATAAATVILGAAAAAALASSASNPFGSGSSGAGSTSQOATIG-----1776
QY 91 ELQMLTNSKTKQDTNQEQPDQAPFQNNGLGTPSADSGGGGCTPDATGGGDPSPATG 150
Db 1777 -----DRDRDRRERERERSGGGGGSGNGNNSNSANSNG--PESAGSG-----SGG 1823
QY 151 GGGDTPTATGGGGSGGCTPTATGCGGCTPTATGCGGCTPTATGCGGCTPTATGCGGCTPTATG 210
Db 1824 GGGSGPAPAGGPNSSGCGGTANSNSGCGG-----GGG-----PALLAGSNSNGV 1869
QY 211 VSDTAGSTEQAKINWVDITIKVAGGEVFDHGATFTADKSMNGDQENQKPMFELAE 270
Db 1870 GSGGAASSNSNSVGGT-----VSGG-----GPGSNQSGSGGGGGP-----ASG 1910
QY 271 ATLKNVNLGENEVDDGIHYKANKNAOEVTI-----DNVHAQNVGEDLTIVKGGGAVTMLN 325
Db 1911 G-----GMGSGAID-----YGGQVAVLTQVANAQAQHYA--ANVAQAAILAASPLG 1955
QY 326 IKNSAKGADKVVQVLANTHLKDINFKADDFG 358
Db 1956 AMESGRKSVHDAQPQIS-----RVEDIKASPGG 1983

RESULT 12
S33799
RNA-binding protein FUS, nuclear - human
N:Alternate names: RNA-binding protein TLS
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
R:Crozac, A.; Aman, P.; Mandahl, N.; Ron, D.
Nature 363, 640-644, 1993
A:Title: Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma.
A:Reference number: S33798; MUID:93288139
A:Accession: S33799
A:Molecule type: mRNA
A:Residues: 1-526 <CRO>
A:Cross-references: GB:S62140; NID:9386156; PIDN:AAB27102.1; PID:9386157
A:Experimental source: liposarcoma
R:Rabbits, T.H.; Forster, A.; Larson, R.; Nathan, P.
Nature Genet. 4, 175-180, 1993
A:Title: Fusion of the dominant negative transcription regulator CHOP with a novel ge
A:Reference number: S36157; MUID:93350637
A:Accession: S36157
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-63, 'S', 66-526 <RAB>
A:Cross-references: EMBL:X71428; NID:9393415; PIDN:CAA50559.1; PID:94210363
A:Experimental source: liposarcoma
C:Genetics:
A:Gene: GDB:FUS
A:Cross-references: GDB:136048; OMIM:137070
A:Map position: 16p11.2-16p11.2
C:Function:
A:Description: RNA binding; probable plays a role in transcriptional regulation
C:Superfamily: RNA-binding protein FMS; ribonucleoprotein repeat homology
C:Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat
F:266-361/Domain: ribonucleoprotein repeat homology <RMR>

```

Query Match 8.7%; Score 191.5; DB 1; Length 526;
 Best Local Similarity 24.2%; Pred. No. 0.00082;
 Matches 96; Conservative 41; Mismatches 148; Indels 111; Gaps 16;

QY 10 QOTTPIDFSLSGSKSPQNT-----PQONTQQAIDPSALLFGSDT 51
 Db 27 QQSSQPIGQGSYSGYSQSTFSGYQSSYSYSGSQNTGTPQGYG--STGYSQSQ 85

QY 52 QKDVAFGTPDSTVONPODASKPNDSQNIAKLISALIMSLOMLTNSKTKQDTNQEQ-- 109
 Db 86 SQSGSYG-----QQSSTFGYGQGPAPPSSTSGSYSSSSQSSSYGQPGQSGYSQDPST 136

QY	110	-----DSQAPFQNNNGSLGTFPSADSGGGGTFPDATGGGGG-----	DTPATGGG	152
Db	137	GGGQOOSTGGQOOSTYPPQGYGQOQNNSSSGGGG-----	GGGGGNNYGODDSSMSGGSG	191
QY	153	G--DTPATGGGGSGGGGTPPATGGSGGTPPATGGGCGVTPQITPOLANPNRTSGTGS	210	
Db	192	GGYGNOPOGGGGGGSGGYGQODRGRGRG-----	SGGGGGG-----GGGG	231
QY	211	VSPDPASTEGQAKTINVKDTIKVGAGVFQGHAKTFPAD-----	KSMGNGDGENOKPMF	265
Db	232	YNRSSGGYEPRGR-----	GGGR--GGRGMGSGDRGKFKGGPRGSGRRHDSQ	279
QY	266	ELAEGATLKNNLGENEYDGIHVAKAKNAOEWTIINV--	HAQNVGEDLITVYGEAGAAVTN	323
Db	280	DNSDNMTIFVYGIGLEN-----	VTIESVADYFKQIG--IKTKTKTGQPMIN	323
QY	324	LNINSSAKGADKRYVOLNANTHLK--	IDNFKADDF	357
Db	324	LYTDRETGKLGAEATVSPDDPSAKAALIDWFQKEF	359	

RESULT 13
 D70807
 hypochlorite-rich glycine-rich protein RV3514 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: D70807
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Gartner, T.; Churcher, C.; Harris, D.; Gordon, S.;
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.;
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: D70807
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1489 <COL>
 A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAAL17751.1; PID:g3252445
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3514
 A:Superfamily: collagen alpha 1(IV) chain

Query Match	8.6%;	Score 189.5;	DB 2;	Length 1489;
Best Local Similarity	29.1%;	Pred. No. 0.0036;		
Matches	71;	Conservative 12;	Mismatches 74;	Indels 87;
				Gaps 11;

QY	118	NGGAGTSAADGGGGGT-----PDATGGGGG-----DTPSATGGGGGP--PPTATGG	161
Dd	729	NAGVAGAGAGSSGAGACTGNSGGAGGTDGGGAGAGACADNPCTGCGTGCGTGGAAGA	788
QY	162	GGSGG---GGT-----PSTATG-----	174
Dd	789	GCAGCAAGTCGTGCMIGTTGNAGVGAGAGCGDGAGAGADADQPcATGCTGFAGGAGG	848
QY	175	-GGSGGPTATG-----GGEGSVTPQITP-----QLANPNR--TSGTGSVDSTASTEBAG	222
Dd	849	AGSGSGSSCACGTNGSGAGGTGCGVVAGAGAGISFSMNSNCGTGTGGTGGDGDGNAG	908
QY	223	KINPVKDTIKYAGEVEFDHGATFTADKSMGNDGOGENOKPMFEIABEATLKNVLGENE	282
Dd	909	-----TGADD--PGKGCTGCTGCTGGSGGAGSGGANFNAGTGTGCTGTGGKGLN	956
QY	283	VDGI 286	
Dd	957	TDDL 960	

RESULT	14
G02127	

fus-like protein - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 23-Jul-1999
 C:Accession: G02127
 R:Itou, K.; Kawase, M.
 submitted to the EMBL Data Library, September 1995
 A:Reference number: G09199
 A:Accession: G02127
 A:Status: preliminary; translated from GR/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-528 <RTO>
 A:Cross-references: EMBL:U35561; NTD:G1040969; P1DN:AAA79948.1; P1D:G1040970
 C:Superfamily: RNA-binding protein EWS:ribonucleoprotein repeat homology
 ;289-364/DomAin: ribonucleoprotein repeat homology <RKM>

Query Match	8.6%;	Score 189;	DB 2;	Length 528;
Best Local Similarity	25.5%;	Pred. No. 0.0011;		
Matches	96;	Conservative 40;	Mismatches 156;	Indels 84;
			Gaps	16

```

QY 10 QQTPLDLFSLKSGSPQPNLT-FGEONTOQAIDPSALLFGSDJTKDYNVFGPDSYVQPO 68
Db 43 QSTDTGYGQSSYSTSYGQSNITGYGTQSAPOCYG-STGYGSSQSSQSSYSTGYQOOS---YPC 98
QY 69 DASKRPNDSQSNIAKLISALIMSLLOMLTNSKKODTNOEQPDSQAPFONNGILG----- 122
Db 99 YGPQPTPSTSGSYGSSSQTSYSGQPPSYGGQPPSYGGQOQSYGGQOQSYNPGYGGQONQYN 158
QY 123 TPSADSGGGG-----TPDATGGGGGDTPSATGGGGGDTPTATGGGGSGGGGTPPTATGG 175
Db 159 SSSSGSGGGGSGYGDQDQSSMSGSGGGG-----GGGGG-----GSGGGGSGYGNDDQSGG 206
QY 176 GSGGTPATGGGEGGYTPQITPOLANPNRTSGTGSYSD-----TAGSTEQGKXINVMYDT 230
Db 207 GGGGGGQDDRRGGRG-----RGRSSSGGGGSGGTYNRSSGGTEPRGR----- 247
QY 231 IKVGAGEYFVDHGATFTAD-----KSMNGGQGEKQKMPFELAEGATLKNVNLGENEYDG 285
Db 248 ---GGGR---GGRGCMGSGSDRGGFNKGFPROGSHHDEQDQNSDNNTLTFYVGLDEN----- 298
QY 286 IHVAKRANAQWYTDNV---HAQNVGEDLTIVKGGGAAYTNLINIKSSAKGADDKVYVOLA 343
Db 299 -----VTIESVADYFKQIG---IKTKNKKTGQPMINLTYTRETGKLGGEATVSPFD 346
QY 344 NTHLK--IDNFKADDF 357
Db 347 PPSAKTAADWFDGKEF 362

```

RESULT 15
 A:70893
 hypothetical glycine-rich protein RV1067c - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
 C:Accession: A70893
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete
 genome number: A70500; MUID:9825987
 A:Accession: A70893
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-667 <COL>
 A:Cross-references: GB:AL021897; GB:AL123456; NID:93256022; PIDN:CAAL17183.1; PID:e125
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: RV1067c
 C:Superfamily: unassigned collagens

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2001, 22:27:05 ; Search time 31.01 Seconds
(without alignments) 468.376 Million cell updates/sec

Title: US-09-597-513-2
Perfect score: 2200

Sequence: 1 MSIGITPPRQOTTPLDFSA.....MTDKKHAIDKTOASTQHTL 424

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	198.5	9.0	1901	YZ08_MYCTU	O53553 mycobacteri
2	193	8.8	2038	FSH_DROME	P13709 drosophila
3	192.5	8.8	401	CSP_PLACG	P08674 plasmodium
4	191.5	8.7	526	FUS_HUMAN	P35637 homo sapien
5	186.5	8.5	622	KICJ_HUMAN	P35527 homo sapien
6	177.5	8.1	1133	N116_YEAST	O02630 saccharomyc
7	177.5	8.1	2329	YS89_CAEEL	O09624 caenorhabdi
8	172.5	7.8	481	LORI_MOUSE	P18155 mus musculu
9	172.5	7.8	518	FUS_MOUSE	P66959 mus musculu
10	172.5	7.8	801	Y747_MYCTU	O53810 mycobacteri
11	171.5	7.7	363	CSP_PLAKH	P02894 plasmodium
12	170	7.7	558	YU83_MYCTU	O10873 mycobacteri
13	170	7.7	778	YU34_MYCTU	P71933 mycobacteri
14	170	7.7	1596	MAM_DROME	P21519 drosophila
15	169.5	7.7	934	MA22_MYCTU	O06794 mycobacteri
16	169	7.7	375	SANT_PLAFV	P09593 plasmodium
17	168	7.6	512	FUS_BOVIN	Q08009 bos taurus
18	168	7.6	543	YD25_MYCTU	O50630 mycobacteri
19	167.5	7.6	603	YD25_MYCTU	O10637 mycobacteri
20	166.5	7.6	287	MSA2_PLAFG	P19260 plasmodium
21	166	7.5	252	GRP1_PHAVU	P10495 phaseolus v
22	165.5	7.5	957	Y278_MYCTU	P16877 mycobacteri
23	164.5	7.5	1317	N145_YEAST	P49687 saccharomyc
24	164.5	7.5	515	Y140_MYCTU	O05594 mycobacteri
25	164.5	7.5	569	KICJ_MOUSE	P02535 mus musculu
26	162.5	7.4	384	GRP1_PENHY	P09789 petunia hyb
27	162.5	7.4	450	SMP1_ENCCU	Q9XZY1 encephalito
28	161.5	7.3	463	YA68_MYCTU	O53416 mycobacteri
29	160.5	7.3	1043	FTF1_DROME	P33244 drosophila
30	159.5	7.2	541	NU57_YEAST	P48837 saccharomyc
31	158.5	7.2	369	VP6_AHSY3	O64909 african hor
32	157.5	7.2	1150	APMU_PIG	P12021 sus scrofa
33	157	7.1	498	Y118_MYCTU	O50615 mycobacteri

34	157	7.1	1210	1	ICEN_PSEFL	P09815 pseudomonas
35	156.5	7.1	161	1	ASFL_HELAN	P22357 heliantus
36	156.5	7.1	174	1	VLFP_MYCHR	O49538 mycoplasma
37	156.5	7.1	645	1	K22E_HUMAN	P35908 homo sapien
38	156.5	7.1	749	1	TROP_HUMAN	O12816 homo sapien
39	156.5	7.1	924	1	IF2_MYCLE	O92519 mycobacteri
40	155.5	7.1	351	1	CSP_PLAKU	P04922 plasmodium
41	155.5	7.1	423	1	BR3A_HUMAN	O01851 homo sapien
42	155.5	7.1	593	1	KICJ_HUMAN	P13645 homo sapien
43	155.5	7.1	672	1	PHX5_MOUSE	P08399 mus musculu
44	155.5	7.1	700	1	NONA_DROME	O04047 drosophila
45	155	7.0	421	1	BR3A_MOUSE	P17208 mus musculu

ALIGNMENTS

```

RESULT 1
ID      YZ08_MYCTU      STANDARD:      PRT: 1901 AA.
AC      O53553;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DE      30-MAY-2000 (Rel. 39, Last annotation update)
DE      HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.
GN      RV3508 OR MY023.15.
OS      Mycobacterium tuberculosis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1773;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=H37RV;
RX      MEDLINE=98295987; PubMed=9634230;
RA      Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA      Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA      Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA      Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA      Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA      Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA      Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA      Taylor K., Whitehead S., Barrett B.G.;
RA      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence."
RL      Nature 393:537-544(1998).
CC      -I- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC      SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
CC      EMBL: AL022022; CAA17745.1; -.
DR      HSSP: P19972; IKVD.
DR      Tuberculist: RV3508; -.
DR      InterPro: IPR000084; -.
DR      Pfam: PF00934; PE; 1.
KW      Hypothetical protein; Repeat; Signal.
FT      SIGNAL
FT      CHAIN
FT      1
FT      30
FT      HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT      RV3508.
SQ      SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;
Query Match 9.0%; Score 198.5; DB 1; Length 1901;
Best Local Similarity 29.2%; Pred. No. 0.0045;
Matches 86; Conservative 13; Mismatches 87; Indels 109; Gaps 14;
119 GGLGTPSADSGGGGTPDATGCGGGGDTPSATGGGGG---DTPPLATG-----GGGSGGG 167

```

```

Db 711 GGGGAGAGNGGCTGSSGAGAGGKGTGGAGSSADNPATGAFGAGGAGCGAAGAG 770
QY 168 GPTATG-GGGGCTPTATG-----GEGGVTPQTTPOLANPNRTSGTG----- 209
Db 771 GAGCATGCTGCTGCVATGAGSAGIGAGGRCGDG-----DGAAGLCTGLSGFD 818
QY 210 ---SVSDTASGTQACGKINNVKDTIKVAGEVDPDGHGATFTADKSMGNGOGENOKPME 266
Db 819 GGGGGGAGGAGGAGAGAGIN-----GAG-----GAG-----GNGDGG----- 849
QY 267 LAEGATLKVNVLGENEVDGTHVAKNAOEVTIDNVAONVGEDLTIVKGGGAAYTNLNI 326
Db 850 -GGAT-GAAGLGDNGCGVGDGAGAGAG-----NCGNAGVGLTAKAGDGAAGNGNG 901
QY 327 KNSAGAGDCKVYVQLNANTHLKIDNFKADDFG-----TWRTNGG 366
Db 902 GAGGAGGAGD-----NFMNGGGGAGGAGGGGGLGAGATTSINANGG 943

```

RESULT 2

FSH_DROME 2 STANDARD: PRT: 2038 AA.

```

AC P13709; P13710;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE FEMALE STERILE HOMEOTIC PROTEIN (FRAGILE-CHORION MEMBRANE PROTEIN).
GN FS(1)H OR FSH.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP MEDLINE-89276730; PubMed-2567251;
RA Haynes S.R., Mozer B.A., Bhatia-Dey N., Dawid I.B.;
RT "The Drosophila fish locus, a maternal effect homeotic gene, encodes
RT apparent membrane proteins.";
RL Dev. Biol. 134:246-257(1989).
CC -1- FUNCTION: REQUIRED MATERIALLY FOR PROPER EXPRESSION OF OTHER
CC HOMEOTIC GENES INVOLVED IN PATTERN FORMATION, SUCH AS UBX.
CC -1- SIMILARITY: HIGH, TO HUMAN RING3 PROTEIN.
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -1- SIMILARITY: CONTAINS 1 ET DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M23221; AAA28540.1; -
DR EMBL: M23222; AAA28541.1; ALT_TERM.
DR EMBL: M15763; AAA70424.1; -
DR EMBL: M15763; AAA70423.1; -
DR EMBL: M15764; AAA70422.1; -
DR PIR: A43742; A43742.
DR HSSP: P04002; 1WFA.
DR FLYBASE: FBgn0004656; fs(1)h.
DR InterPro: IPR001487; -.
DR Pfam: PF00439; bromodomain; 2.
DR PRINTS: PRO0503; BROMODOMAIN.
DR PROSITE: PS00633; BROMODOMAIN_1; 2.
DR PROSITE: PS00633; BROMODOMAIN_2; 2.
KW Developmental protein; Bromodomain; Transmembrane; Repeat.
FT DOMAIN 51 123 BROMODOMAIN 1.
FT DOMAIN 495 567 BROMODOMAIN 2.
FT DOMAIN 945 1106 ET DOMAIN.
FT TRANSMEM 330 350 POTENTIAL.

```

```

FT TRANSMEM 451 471 POTENTIAL.
FT TRANSMEM 750 770 POTENTIAL.
FT TRANSMEM 790 810 POTENTIAL.
FT TRANSMEM 816 830 POTENTIAL.
FT TRANSMEM 874 894 POTENTIAL.
FT TRANSMEM 1731 1751 POTENTIAL.
FT TRANSMEM 1939 1959 POTENTIAL.
FT VARIANT 909 909 G -> A.
FT VARIANT 1022 1022 H -> RPKY.
SQ SEQUENCE 2038 AA; 205332 MW; 849E0706D50A0098 CRC64;

```

Query Match 8.8%; Score 193; DB 1; Length 2038;

Best local Similarity 24.0%; Pred. No. 0.0092;

Matches 80; Conservative 40; Mismatches 133; Indels 80; Gaps 14;

```

QY 36 TQQAIDPSALIFSGDTQKDVNFGTP-----DSTVONPQDASKFNDOSNTAKILSLIMS 90
Db 1721 TSAVAQAAAIATAATATAAATLGAATAAALASSANPSGSSSGAGSTSQATITG---- 1776
QY 91 LLOMLTNSNKKOPTNQEQPQAPFQNNGLGTPPSADSGSGTPTATGGGSDTPTATG 150
Db 1777 ---DNRDRERERERRSSGGGSGGNGNNSNSANSNG---PGSAGSG-----SGG 1823
QY 151 GGGDTPTATGGGSGGCTPTATGGSGGCTPTATGGGEGGVTPQTTPOLANPNRTSGTG 210
Db 1824 GGGSGPASAGAPNPSGGCTANSNGGGG-----GGG-----PALNAGSNNSGV 1869
QY 211 VSPTAGSTEDGAKINNVKDTIKVAGEVDPDGHGATFTADKSMGNGOGENOKPMEFLAEG 270
Db 1870 GSGGAASSNSNSVSGT-----VGS-----PGSNSGSSSGGGGGP-----ASG 1910
QY 271 ATLKNVNLGENEVDGTHVAKNAOEVTI-----DNYHAONVGEDLTIVKGGGAAYTNLN 325
Db 1911 G-----GMSGALD-----YGGVAVLTQVANAQAQHYA---AAYAAQALIASPLG 1955
QY 326 IKNSAGAGDCKVYVQLNANTHLKIDNFKADDFG 358
Db 1956 AMESGRKSVHDAQPQIS-----RVEDIKASPG 1983

```

RESULT 3

CSP_PLACG 3 STANDARD: PRT: 401 AA.

```

AC P08674;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).
OS Plasmodium cynomolgi (strain Gombak).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5830;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-87102878; PubMed-3802196;
RA Galinski M.R., Arnot D.E., Cochrane A.H., Barnwell J.W.,
RA Nussenzweig R.S., Enea V.;
RT "The circumsporozoite gene of the Plasmodium cynomolgi complex.";
RL Cell 48:311-319(1987).
CC -1- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC VERTEBRATE HOST).
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M15100; AAA29536.1; -
 DR PIR: E26255; 02ZQAC.
 DR InterPro: IPR000884; -
 DR InterPro: IPR003067; -
 DR Pfam: PF00090; tSP.1; 1.
 DR PRINTS: PRO1303; CRCMSPRZOTE.
 DR KMW: Malaria; Sporozoit; Repeat; Signal.
 DR SGNL: 1 19 PROBABLE.
 FT CHAIN 20 401 CIRCUMPOROZITE PROTEIN.
 FT DOMAIN 98 278 17 X 11 AA TANDEM REPEATS OF [DG]-G-A-A-
 FT A-A-G-G-G-G-N.
 FT A-A-G-G-G-G-N.
 SQ SEQUENCE 401 AA; 36664 MW; 57d66268238503E CRC64;
 Query Match 8.8%; Score 192.5; DB 1; Length 401;
 Best local similarity 26.1%; Pred. No. 0.0018;
 Matches 90; Conservative 27; Mismatches 121; Indels 107; Gaps 15;
 QY 65 QNPQD--ASRPNDQSNIATLISALIMSLQMLTNSNRKQDTNOEPDSQAPQNN----118
 DB 64 ENPKDEGADKP-----KKRDEKQVEP--KKPRENKLKOP 95
 QY 119 -----GGLTGPSADSGGCGTTPD---ATGGGGGDPSPATGGGGGCTPPTATGGGS---164
 DB 96 REDGAAAAAGGGNDGAAAAAGGGGNDGAAAAAGGGGNDGAAAAAGGGGNDGAA 155
 QY 165 -----GGG--GTPTATGGGSGGTPTATGGGEGGYTPPTPOLANPNRFGSGVSVDTA---215
 DB 156 AAGGGGNDGAAAAAGGGGNDGAAAAAGGGGNDGAA--AAGGGGNDGAAAAAGGGGNDGAAAA 213
 QY 216 -----GSTEQAQKINNVKDTIKVGAGEVFDGHATFTADKSGND--256
 DB 214 GGGGNGGAAAAAGGGGNDGAAAAAGGGGNDGAAAAAGGGGNDGAAAAAGGGGNDGAA 273
 QY 257 -----QSENRKPMELAEKATIKVNVNGENEVDGIHKARKNAQEVITIDNVAQVGE 308
 DB 274 AGGNNARADQDP-----AGGNKKAEGAGNAGAG--QAQNNFANVPN--AKLYKE 323
 QY 309 DLTVKGGGGAATVNLINIKNSAKG-----ADDKVYQVLNAN 344
 DB 324 YLDKIRSTLGVEMSPCTV--TCGKGVMRKRVSAANKPELDAN 366
 RESULT 4
 FUS_HUMAN ID FUS_HUMAN STANDARD; PRT; 526 AA.
 AC P35637;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE RNA-BINDING PROTEIN FUS (ONCOGENE TLS) (TRANSLATED IN
 DE LIPOSARCOMA PROTEIN) (POMP75) (75 KDA DNA-PAIRING PROTEIN).
 GN FUS OR TLS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RX MEDLINE=93288139; PubMed=8510758;
 RA Crozat A., Aman P., Mandahl N., Ron D.;
 RT Fusion of CHOP to a novel RNA-binding protein in human myxoid
 RT liposarcoma."
 RL Nature 363:640-644(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RX MEDLINE=93350637; PubMed=7503811;
 RA Rabbitts T.H., Forster A., Larson R., Nathan P.;
 RT Fusion of the dominant negative transcription regulator CHOP with a
 RT novel gene FUS by translocation t(12;16) in malignant liposarcoma.";

RL Nat. Genet. 4:175-180(1993).
 RN [3]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RX MEDLINE=99013873; PubMed=9795213;
 RA Morohoshi F., Ootsuka Y., Arai K., Ichikawa H., Mitani S.,
 RA Munakata N., Ohki M.;
 RT "Genomic structure of the human RBP56/MTAFI68 and FUS/TLS genes.";
 RL Gene 221:191-198(1998).
 RN [4]
 RP SEQUENCE OF 235-244; 307-312; 335-345 AND 349-357, & CHARACTERIZATION.
 RX MEDLINE=99369251; PubMed=10567410;
 RA Beechold H., Kuroda M., Sok J., Ron D., Lopez B.S., Akhmedov A.T.;
 RT "Human 75-kDa DNA-pairing protein is identical to the pro-oncoprotein
 RT TLS/FUS and is able to promote D-loop formation.";
 RL J. Biol. Chem. 274:34337-34342(1999).
 RN [5]
 RP SEQUENCE OF 265-276; 317-331 AND 335-357, AND IDENTIFICATION.
 RX MEDLINE=94243799; PubMed=8187069;
 RA Ichikawa H., Shimizu K., Hayashi Y., Ohki M.;
 RT "An RNA-binding protein gene, TLS/FUS, is fused to ERG in human
 RT myeloid leukemia with t(16;21) chromosomal translocation.";
 RL Cancer Res. 54:2865-2868(1994).
 CC -1- FUNCTION: BINDS BOTH SINGLE-STRANDED AND DOUBLE-STRANDED DNA AND
 CC PROMOTES ATP-INDEPENDENT ANNEALING OF COMPLEMENTARY SINGLE-
 CC STRANDED DNAs AND D-LOOP FORMATION IN SUPERHELICAL DOUBLE-STRANDED
 CC DNA. MAY PLAY A ROLE IN MAINTENANCE OF GENOMIC INTEGRITY.
 CC -1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG ISOFORM (SHOWN HERE)
 CC AND A SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBQUITOUS.
 CC -1- DISEASE: A FORM OF MALIGNANT MYXOID LIPOSARCOMA IS CHARACTERIZED
 CC BY A CHROMOSOMAL TRANSLOCATION T(12;16)(p11;p11) THAT INVOLVES FUS
 CC AND CHOP.
 CC -1- DISEASE: A FORM OF ACUTE MYELOID LEUKEMIA (AML) IS CHARACTERIZED
 CC BY A CHROMOSOMAL TRANSLOCATION T(16;21)(p11;q22) THAT INVOLVES
 CC FUS AND ERG.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 RANBP-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S62140; AAB27102.1; -
 DR EMBL: S62138; AAB27103.1; ALT_SEQ.
 DR EMBL: X71427; CAA50558.1; ALT_SEQ.
 DR EMBL: X71428; CAA50559.1; ALT_SEQ.
 DR EMBL: AF071213; AAC35285.1; -
 DR EMBL: AF071213; AAC35284.1; -
 DR HSSP: P09651; 1HA1.
 DR MIM: 137070; -
 DR MIM: 151900; -
 DR InterPro: IPR000504; -
 DR InterPro: IPR001876; -
 DR Pfam: PF00076; rim; 1.
 DR Pfam: PF00641; zf-RanBP; 1.
 DR PROSITE: PS50102; RRM; 1.
 DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.

FT VARIANT 167 167 L -> S (IN EPPK).
 FT BEST LOCAL SIMILARITY 23.8%; Pred. No. 0.0057; DB 1; Length 622;
 FT CONFLICT 12 12 T -> SR (IN CAAB2315).
 SQ SEQUENCE 622 AA; 61987 MW; 898C3825D4B5ED94 CRC64;

Query Match 8.5%; Score 186.5; DB 1; Length 622;
 Best Local Similarity 23.8%; Pred. No. 0.0057; DB 1; Length 622;
 Matches 88; Conservative 50; Mismatches 131; Indels 101; Gaps 16;

QY 119 GGLG-----TPSADSGGGGTPDATGGGGGDPSPATGGGGGDPPTATGGGGSGG 166
 DB 19 GGLGSGGSSINSYSRFGSSSGRGGRSSSSGCGSSSVCGGSGGSGSGYSYGGSGG 78
 QY 167 GGPPTATGGSGGTPPTATGGGGGVPPTPOLANPRTSGTGSVSTPASTGAGKINV 226
 DB 79 GFSASSIGGFGGSGRFGGASG-----GYSSSGGFGGFGG----- 116
 QY 227 VKPTIKVAGEVF-DGHAFTADKSMNGDGENCKPMFELAGATLKNVNLGENEVDG 285
 DB 117 -----GSGGGFGGSGGSGGFGGAGGCGGGLTANKSTMOELN--SRLAS 165
 QY 286 IHAKAKNAQVTTIDNVHAQVNGEDLITVKEGGAAYTNLKNSSAGADKVVOL--N 342
 DB 166 YLDKVALLEANNND--LEKKIDWYDKKP-AAIOKNYSPIYTTIDLDKQIVDLVGN 221
 QY 343 ANTHLKDNRK--ADDF-----GTWRTNGSKQ-FDDMSTELNGIEANHGK-- 385
 DB 222 NKTLLDIDNRMFLDPRFKFEMEQNLROGVADINGLROYLDLITMEKSDLEMQYETLQ 281
 QY 386 ---FALVKSDDDLKLATG-----NIA-----MTDVKHAYD-----K 414
 DB 282 EELMALKKHKEKESQLTGONGSDVNEIVNAEGKDLTKLNDMRQETOLIAKRRDIE 341
 QY 415 TOASTOTTEL 424
 DB 342 NOYETOTIOTI 351

RESULT 6
 N116 YEAST STANDARD; PRT; 1113 AA.
 ID N116 YEAST STANDARD; PRT; 1113 AA.
 AC 002630;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE NUCLEOPORIN NUP116/NSP16 (NUCLEOPORIN PORE PROTEIN NUP116/NSP16).
 GN NUP116 OR NSP16 OR YMR047C OR YMR9532.12C.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054906; PubMed=1385442;
 RA Mente S.R., Rout M.P., Blobel G.;
 RT "A new family of yeast nuclear pore complex proteins";
 RL J. Cell Biol. 119:705-723(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93099880; PubMed=1464327;
 RA Wimmer C., Doye V., Grandi P., Nehrbass U., Hurt E.C.;
 RT "A new subclass of nucleoporins that functionally interact with
 nuclear pore protein NSP1";
 RL EMBO J. 11:5051-5061(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
 CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
 CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.

CC -1- SUBUNIT: INTERACTS WITH KAP95.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
 CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
 CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z15036; CA78754.1; -
 DR EMBL: X68108; CA48228.1; -
 DR EMBL: Z48502; CA488413.1; -
 DR PIR: S28537; S28537.
 DR PIR: S28925; S28925.
 DR PIR: A44402; A44402.
 DR SGD: S0004650; NUP116.
 KW Nuclear protein; Transport; Repeat.
 FT DOMAIN 205 715
 FT 37 X 6 AA APPROXIMATE REPEATS OF
 FT G-L-F-G.
 FT CONFLICT 26 26 G -> A (IN REF. 1).
 FT CONFLICT 536 536 S -> G (IN REF. 1).
 FT CONFLICT 720 720 S -> P (IN REF. 1).
 FT CONFLICT 1018 1018 S -> Y (IN REF. 1).
 FT CONFLICT 1023 1023 I -> Y (IN REF. 1).
 SQ SEQUENCE 1113 AA; 116234 MW; FBA0B9A9EA958213 CRC64;

Query Match 8.1%; Score 177.5; DB 1; Length 1113;
 Best Local Similarity 22.1%; Pred. No. 0.029; DB 1; Length 1113;
 Matches 107; Conservative 59; Mismatches 182; Indels 137; Gaps 24;

QY 31 FGEQNTQOAIIDPSALIFGSDTKQDVNFGTDPSTVQNPQD-----ASKPNDQSNIARLIS 85
 DB 484 FGGQNNQSQSGPLFGQTQNNQNPFG--QNGLOQPOQNNSLFGAPRTGNGT----- 535
 QY 86 ALINSLQMLTNSNKKQDT-----NOEOPDSQAPQON-----NGGL--GTPSADSGG-- 131
 DB 536 -----SLFSNSTTQNSGICGNNLQOQSGGLFQNKQOPASGGLFGSKPSNVTGGLFG 588
 QY 132 -----GTPDATGGGGGDPSPATG---GGGGDP--PRATGG--GSGGGTPPTATGGGG 178
 DB 589 NMQVANQNNPASTISGGLFGSKPATGSLFGTSTAPNASSGGLFGSNNASNTAATTNSTG 648
 QY 179 -----GTPATGGGEGGVTPQITPOLANPRTSGTGSVSTPASTGEOAGKINVKD 229
 DB 649 LFGKKPYGAGASTSAGGLF-----NNNSLANSNGSTGLFGSNNTSQS 693
 QY 230 TIKVAGEVFDFGAGTETA-----DKSNGND-----QGENCKPMFEL 267
 DB 694 T---NAGGLFQNNSTSTSGGLFGSPQSPQSQNALQOQOQOQRIQIQQNNPYGTNEL 750
 QY 268 AEGATLKNVNLGENEVGCIHYAKNAQEVTTIDNVHAQVNGEDLITVYGEAGTAATNINIK 327
 DB 751 FSKATVTNTVSYPIQSPATRKADERRKASLTNAV-KWIPPTLETA-----LKTNSVM 804
 QY 328 NSSAKGADKV-VOLN-ANTHLKIDN-----FKAD--DFGTWVR--- 363
 DB 805 DKQIKYDPLKLISTIDKKNQOIALSNQOENLDESILKASLELFPNPKRSKKNLNNRKM 864
 QY 364 -----NGKQDFDMSIELNGIEANHGKALVKSDDDLKLATGNIAMTDVKAHYDKTQ 416
 DB 865 LIASEEKNNGSQNNDMFKSKSEBO---ETILGKPKKDEKTANGCBRMV-ISSKNDGED 920
 QY 417 ASTQH 421
 DB 921 SATKH 925

RESULT 7

```

YS89_CAEEL STANDARD: PRT: 2329 AA.
ID YS89_CAEEL
AC 009624:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHELTICAL 254.3 KDA PROTEIN ZK945.9 IN CHROMOSOME II.
GN ZK945.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodetrinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilkinson-Sproat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBP databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z48544; CAA8442.1;
DR WormPep: ZK945.9; CE01740.
DR InterPro: IPR000203;
DR InterPro: IPR001024;
DR Pfam: PF01825; GPS; 1.
DR Pfam: PF01477; PLAT; 1.
KM Hypothetical protein: Repeat: Transmembrane.
FT DOMAIN 22 115 SER/THR-RICH.
FT TRANSMEM 250 389 GY/SER-RICH.
FT TRANSMEM 577 577 POTENTIAL.
FT TRANSMEM 606 626 POTENTIAL.
FT TRANSMEM 1162 1182 POTENTIAL.
FT TRANSMEM 1290 1310 POTENTIAL.
FT TRANSMEM 1467 1487 POTENTIAL.
FT TRANSMEM 1498 1518 POTENTIAL.
FT TRANSMEM 1541 1561 POTENTIAL.
FT TRANSMEM 1602 1622 POTENTIAL.
FT TRANSMEM 1637 1657 POTENTIAL.
FT TRANSMEM 1717 1737 POTENTIAL.
FT TRANSMEM 1935 1955 POTENTIAL.
FT TRANSMEM 1990 2010 POTENTIAL.
FT TRANSMEM 2039 2059 POTENTIAL.
FT TRANSMEM 2088 2108 POTENTIAL.
FT TRANSMEM 2134 2154 POTENTIAL.
FT TRANSMEM 2189 2209 POTENTIAL.
SQ SEQUENCE 2329 AA; 254348 MW; 4D9D423060F0CAB5 CRC64;

```

```

Query Match 8.1%; Score 177.5; DB 1; Length 2329;
Best Local Similarity 22.7%; Pred. No. 0.063;
Matches 96; Conservative 52; Mismatches 110; Indels 165; Gaps 19;

125 SADSGL--GCTPATGGGGGDPATGCGGDDPPTATG-----GGGSG 165
      ||: || ||: || ||: || ||: || ||: || ||: || ||: ||
252 SDDAGKKTGGT-GATGGTGG-----TSGSGSATLTSTGDAVRSTTSGSGSGSGAG 305
      ||: || ||: || ||: || ||: || ||: || ||: || ||: ||
166 GGGTPATGGGSGTPTATGGGEGVTPTPTPOLANPNRTSGT-----GS 210
      ||: || ||: || ||: || ||: || ||: || ||: || ||: ||
306 GSGT-TASGSGSG-----SSGTGSDGVNSKTTALNDGSGGATTPGSHLGGGSGSGS 361
      ||: || ||: || ||: || ||: || ||: || ||: || ||: ||
211 VSDTAGSTEQACKINVKDTIKYAGEVEVDGSGAFTADKSMKNGOGENQKPMFELAE 270
      ||: || ||: || ||: || ||: || ||: || ||: || ||: ||
362 GSDSNSSGSGVSTKSSGSDP-----SSGSDSGGANGAFSATAGSTRTTKT 407
      ||: || ||: || ||: || ||: || ||: || ||: || ||: ||
271 ----ATLKVNNGE-----NEWDT-----HYAKAKAQEVTTIDNVH----- 302
      ||: || ||: || ||: || ||: || ||: || ||: || ||: ||

RESULT 8
LORI_MOUSE STANDARD: PRT: 481 AA.
ID LORI_MOUSE
AC P18165;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE LORICRIN.
GN LOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90275605; PubMed-2190691;
RA Mehrel T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.,
RA Cheng C., Lichti U., Bisher M.E., Steven A.C., Steinat P.M.,
RA Yuspa S.H., Koop D.R.;
RT "Identification of a major keratinocyte cell envelope protein,
RT loricrin."
RL Cell 61:1103-1112(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BAB/C;
RX MEDLINE-95556248; PubMed-7738016;
RA Diserio D., Jones A., Longley M.A., Bundman D., Rothnagel J.A.,
RA Koop D.R.;
RT "The proximal promoter of the mouse loricrin gene contains a
RT functional AP-1 element and directs keratinocyte-specific but not
RT differential expression."
RL J. Biol. Chem. 270:10792-10799(1995).
CC -1- FUNCTION: MAJOR KERATINOCYTE CELL ENVELOPE PROTEIN.
CC -1- SUBUNIT: MONOMERS ARE CROSSLINKED BY DISULFIDE AND N-(GAMMA-
CC GLUTAMYL) LYSINE ISODIPEPTIDE BONDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M34398; AAA39444.1;
DR EMBL: U09189; AAA82152.1;
DR PIR: A35628; A35628.
DR HSSP: P10968; 1WGC.
DR MGD: MGI:96816; LOR.
KM keratinocyte.
SQ SEQUENCE 481 AA; 37830 MW; 97349A786FF239FE CRC64;

```

Query Match 7.8%; Score 172.5; DB 1; Length 481;


```

Best Local Similarity 32.5%; Pred. No. 0.022;
Matches 51; Conservative 15; Mismatches 78; Indels 13; Gaps 3;

OY 118 NGGLCTPSADSG--GGTPTATGCGGDDTPSANTGCGGDDPTTAGCGSGGTPTATG 174
   :|::|||:|||||:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 114 SGGCGSSCGGGSYSCGGSSCGGGSGGGVXYSGGGGG-SSCGGSSGGGGSSCG 172
OY 175 GSGGTPRATPGEGEGVYRPLITPOLANNRFTSGTSSVDIAGTDQAKINVKRT---- 230
   |||||:|:|||||:|:|||||:|:|||||:|:|||||:|:|||||:|:
Db 173 GSSGGGSGYSCGGSSCGGGSGGGGKRYSGGGSSCGGGSGGGSSCGG 232
OY 231 -----IKVAGEVFDPGHGATFADRSKMNGDGENOK 262
   |::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 233 SGGGSSCGGGGGGYSGGGGTSCGGSSSGGGGGSSSQ 269

RESULT 9
FUS_MOUSE STANDARD: PRT: 518 AA.
AC P56959;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-BINDING PROTEIN FUS (PIGPEN PROTEIN).
GN FUS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN RNP [1]
RP SEQUENCE FROM N.A.
RA Alappat S.R., Zhang M., Zhao X., Alliegro M.A., Alliegro M.C.,
RA Burdesh C.A.;
RT "Regulation of piipen expression in mouse embryos.";
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: BINDS BOTH SINGLE-STRADED AND DOUBLE-STRADED DNA AND
CC PROMOTES ATP-INDEPENDENT ANNEALING OF COMPLEMENTARY SINGLE-
CC STRADED DNAs AND D-LOOP FORMATION IN SUPERHELICAL DOUBLE-STRADED
CC DNA. MAY PLAY A ROLE IN MAINTENANCE OF GENOMIC INTEGRITY (BY
CC SIMILARITY).
CC -1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics institute. There are no restrictions on lists
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
Cc CC
DR EMBL; AF224264; AAF70602.1; -.
DR MGD; MG1:1353633; FUS.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; FALSE NEG.
KW RNA-binding; DNA-binding; Nuclear protein; Repeat; Zinc-finger; Zinc;
KW Metal-binding.
FT DOMAIN 1 167 GLN/GLY/SER/TYR-RICH.
FT DOMAIN 168 265 GLY-RICH.
FT DOMAIN 278 364 RNA-BINDING (RRM).
FT DOMAIN 364 518 ARG/GLY-RICH.
FT ZN-FING 421 440 C4-TYPE (POTENTIAL).
SQ SEQUENCE 518 AA; 52673 MW; E06F231BREED7806 CNC64;

Query Match 7.8%; Score 172.5; DB 1; Length 518;
Best Local Similarity 23.7%; Pred. No. 0.024;
Matches 84; Conservative 45; Mismatches 122; Indels 103; Gaps 16;

22 SGKSQPNTFEQNT-----QAIDPSALLFGSDQTQDVNFETPDSTVQNFDQASKPND 75

```

[illegible]

RESULT	10
Y747_MYCTU	
ID	Y747_MYCTU
AC	053810;
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	30-MAY-2000 (Rel. 39, Last annotation update)
DE	HYPOTHEITICAL PE-PGRS FAMILY PROTEIN RV0747 PRECURSOR.
GN	RV0747 OR MYV041.21.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX	NCBI_TaxId=1773;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=H37RV.
RX	MEDLINE=96295987; PubMed=9634230;
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA	Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Hornsby T., Jagels K., Krogh A., McLean A., Moule S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA	Taylor K., Whitehead S., Barrall B.G.;
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RT	complete genome sequence."
RL	Nature 393:537-544(1998).
CC	-I- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC	SUBFAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and that statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AL021958; CAA17514.1; -.
DR	TubercuList; RV0747; -.
DR	InterPro; IPR000084; -.
DR	Pfam; PF00934; Pf; 1.
FT	Hypothetical protein; Repeat; Signal.
FT	SIGNAL 1 30
FT	CHAIN 31 801
FT	POTENTIAL.
FT	HYPOTHEITICAL PE-PGRS FAMILY PROTEIN
FT	RV0747.
FT	SEQUENCE 801 AA; 65407 MW; EA54C9BF45A00F41 CRC64;


```

CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE
CC OBVIOUSLY LOCATED. DURING LATER STAGES THEY ACCUMULATE IN THE
CC CENTRAL NERVOUS SYSTEM.
CC -1- MISCELLANEOUS: THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS. 21
CC POLY-GLN RUNS (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10
CC AA), 3 POLY-ASN (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5
CC AA) RUNS.
CC -1- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
CC YEAST AND MAMMALIAN REGULATORY PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X54251; CAA38152.1; -.
CC PIR: A33106; A33106.
CC PIR: A36391; A36391.
CC FLYBASE: FBgn0002643; mam.
CC Neurogenesis; Nuclear protein; Repeat.
CC Gln-Rich.
CC Gln-Rich.
CC ARG/LYS-RICH (BASIC).
CC Gln-Rich.
CC ASN-RICH.
CC GLY/ASN-RICH.
CC GLN-RICH.
CC FT DOMAIN 127 190
CC FT DOMAIN 166 219
CC FT DOMAIN 259 304
CC FT DOMAIN 355 388
CC FT DOMAIN 392 406
CC FT DOMAIN 407 440
CC FT DOMAIN 651 671
CC FT DOMAIN 700 714
CC FT DOMAIN 759 816
CC FT DOMAIN 987 996
CC FT DOMAIN 1060 1079
CC FT DOMAIN 1092 1107
CC FT DOMAIN 1237 1252
CC FT DOMAIN 1492 1496
CC FT DOMAIN 1559 1592
CC FT DOMAIN 1596 AA; 167717 MW; B944D86EF35D605 CMC64;
CC SEQUENCE

```

```

Query Match 7.7%; Score 170; DB 1; Length 1596;
Best Local Similarity 22.7%; Pred. No. 0.1;
Matches 76; Conservative 37; Mismatches 100; Indels 122; Gaps 15;

```

```

QY 48 GSDTQKDVNFGTDPSTVONPDASKPNDOSNIAKLIALIMSL-----IQMLT 96
DB 281 GSNMTGNNTN--NNGNSTNN--GSSNNGSENLTKEFVEIYVQOLEFTTSPANSQPOIST 336
QY 97 NSKKKQDTNDEQDQAPFONNGILGTPSADSGGGTPTATGCGGDTPTATGCGGDTPT 156
DB 337 NNTVYKALTN-----TSVKSPEVGG-----GGGG--GSGNNNNNGG-- 375
QY 157 TATGGGSG-----GGGT-----PTATGGGSGGTPATGGG 187
DB 376 ---GGGGNGNNNGNDHQQQQQQHQQQQQQGGGLGGLGNNRGSGGGMATGPBGV 432
QY 188 EGGVTPQ--ITPOLANPNRISGTSVSDTAGSTED-----AGKINYVKPTIKYAGEVF 239
DB 433 AGGILGGMGMPNMAASQAOKSALGNLNLVCKRPPDHDPDLGLD-----KGGGGGQF 486
QY 240 DGHGATFTLAKSGNGND-----QGENOKPMFELAEAGATLKNNVLGSENV 283
DB 487 PGDPDLIGDENSENDTFFKLINNLADFNPSFLDGPDEKFLDL-----KTE 533
QY 284 DGIIHVAKNAQEVYTIQVHAQNVGEGLITVKGEGG 318
DB 534 DGIKVEPPNADLL-----NSLANKVSGG 557

```

RESULT 15

```

WA22 MYCTU
ID WA22 MYCTU STANDARD: PRT: 914 AA.
AC 006794;
DT 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE WAG22 ANTIGEN PRECURSOR.
GN WAG22 OR RV1759C OR MWCY28.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98293967; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigleier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLaren J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellern S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RA Taylor K., Whitehead S., Barrett B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.
RT Nature 393:537-544 (1998).
RT Nature 393:537-544 (1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z95890; CAB09322.1; -.
CC HSSP: P41140; 2SFA.
CC Tuberculin; RV1759C; -.
CC Interp: IPR000084; -.
CC Pfam: PF00934; PE; 1.
CC Antigen; Repeat; Signal.
CC SIGNAL 1 30
CC CHAIN 31 914
CC POTENTIAL.
CC WAG22 ANTIGEN.
CC SEQUENCE 914 AA; 74354 MW; F6953C3DBBE66A8 CRC64;

```

```

Query Match 7.7%; Score 169.5; DB 1; Length 914;
Best Local Similarity 28.4%; Pred. No. 0.06;
Matches 67; Conservative 17; Mismatches 95; Indels 57; Gaps 9;

```

```

QY 20 ALSGKSPQPTFGEQNTQQAIDPSALLFGSDTQKDVNFGTDPSTVONPDASKPNDOSN 79
DB 479 ALGKS-----GFG-----GEGGLLGD-----GYNAPEST-----SPWHN 510
QY 80 IAKLIALIMSLIOMLTNSKKQDTNDEQDQAPFONNGILGTPSA--DSGGGTPDAT 137
DB 511 LQDDILSFINEPTREALTG-----RPLIGNDSSTPGGDDGAGCMIFGN 555
QY 138 GGGGDTPTATGCGGDTPTATGCGGSGGGTPTATGCGGSGGTPATGCGGSGGTPAT 197
DB 556 GGGGAGAAAGTNGSAG-----GAGGAGILFTGTGAGAGAGVGTAGAGAGAGAG--SA 607
QY 198 QLANPNTSGTSVSDTAGSTEDQAKINVKDITIKYAGAGVFPDGHATFTADSMG 253
DB 608 FLIGSGGTGVGGAATTTGGVGAGG---MAGLLIGMAGLGGCGGGAFTAGVTTG 659

```

Search completed: October 4, 2001, 22:33:52
Job time: 407 sec

Fri Oct 5 10:04:31 2001

us-09-597-513-2.rsp

Page 11

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2001, 22:26:34 ; Search time 60.34 Seconds
(without alignments)
929.687 Million cell updates/sec

Title: US-09-597-513-2

Perfect score: 2200
Sequence: 1 MSIGITPPRQQTITPLDPSA.....MTDKAHAYDKTQASTQHTL 424

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP unclassified:*
13: SP vertebrate:*
14: SP virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2200	100.0	424	2 087264	087264 pseudomonas
2	1591.5	72.3	441	2 087327	087327 pseudomonas
3	563	25.6	447	2 091AW2	091AW2 erwina amy
4	559	25.4	447	2 054508	054508 erwina amy
5	318	14.5	224	2 09RHM0	09RHM0 bacillus sp
6	315	14.3	222	2 09X622	09X622 bacillus su
7	291.5	13.2	221	2 034310	034310 bacillus su
8	239	10.9	266	2 09EX16	09EX16 streptomyce
9	216	9.8	3190	5 001368	001368 drosophila
10	215	9.8	730	10 09ZU23	09ZU23 arabidopsis
11	214	9.7	3275	5 09W321	09W321 drosophila
12	208	9.5	215	3 000843	000843 necria hae
13	204.5	9.3	680	10 P93658	P93658 brassica na
14	203.5	9.2	233	3 000845	000845 necria hae
15	196.5	8.9	1226	5 021835	021835 caenorhabd1
16	194	8.8	3016	2 P73590	P73590 synecocyst
17	193	8.8	1937	5 09W3L3	09W3L3 drosophila
18	189.5	8.6	620	5 09V7U0	09V7U0 drosophila
19	189.5	8.6	1489	2 053559	053559 mycobacteri

20	189	8.6	528	4 013344	013344 homo sapien
21	188.5	8.6	705	10 004310	004310 arabidopsis
22	188	8.5	667	2 053415	053415 mycobacteri
23	187.5	8.5	813	2 050279	050279 mycoplasma
24	186.5	8.5	240	3 093877	093877 fusarium ox
25	186.5	8.5	698	14 09J3U0	09J3U0 ectocarpus
26	185.5	8.4	1079	2 053557	053557 mycobacteri
27	185	8.4	207	10 043522	043522 lycopersico
28	185	8.4	347	5 09GZC7	09GZC7 trypanosoma
29	184.5	8.4	697	5 09GRW7	09GRW7 drosophila
30	183	8.3	731	2 050415	050415 mycobacteri
31	182.5	8.3	1538	2 053395	053395 mycobacteri
32	181.5	8.2	1381	2 053552	053552 mycobacteri
33	180.5	8.2	348	13 093397	093397 cyprinus ca
34	179.5	8.2	694	2 053212	053212 mycobacteri
35	179.5	8.2	941	2 09XD54	09XD54 moraxella c
36	178.5	8.1	242	3 004701	004701 fusarium so
37	178.5	8.1	741	2 006808	006808 mycobacteri
38	178	8.1	770	2 085783	085783 myxococcus
39	177.5	8.1	388	5 094675	094675 plasmodium
40	177.5	8.1	491	2 006818	006818 mycobacteri
41	177	8.0	912	2 09XD52	09XD52 moraxella c
42	177	8.0	944	5 09VK59	09VK59 drosophila
43	176.5	8.0	1963	2 09XC03	09XC03 salmonella
44	176	8.0	447	10 09S1A8	09S1A8 arabidopsis
45	176	8.0	618	2 006801	006801 mycobacteri

ALIGNMENTS

```

RESULT 1
ID 087264 PRELIMINARY; PRT; 424 AA.
AC 087264;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE HRPW.
GN HRPW.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=98422476; PubMed=9748456;
RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA Collier A.;
RT "The Pseudomonas syringae pv. tomato HRPW protein has domains similar
RT to harpins and peptate lyases and can elicit the plant hypersensitive
RT response and bind to peptate."
RT J. Bacteriol. 180:5211-5217(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=20243785; PubMed=10781092;
RA Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
RA Penlocki-Ocwleja T., Van Dijk K., Collier A.;
RT "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
RT mosaic structure composed of a cluster of type III secretion genes
RT bounded by exchangeable effector and conserved effector loci that
RT contribute to parasitic fitness and pathogenicity in plants."
RT Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
DR EMBL: AF005221; AAC62526.1; -.
DR EMBL: AF320065; AAF71503.1; -.
SQ SEQUENCE 424 AA; 42910 MW; 7AF4ED059BE79D2E CRC64;

```

Query Match 100.0%; Score 2200; DB 2; Length 424;
Best Local Similarity 100.0%; Pred. No. 1.5e-128;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MSIGITPRPOQTTPPLDPSALSGKSPQNTFGEONTQOALDPSALLFGSDTOKVNGFP 60
DB 1 MSIGITPRPOQTTPPLDPSALSGKSPQNTFGEONTQOALDPSALLFGSDTOKVNGFP 60
QY 61 DSTVONPODASRPNDQSNIAKLISALIMSLQMLTNSNRKQDPTNOQPDSCAPQONNG 120
DB 61 DSTVONPODASRPNDQSNIAKLISALIMSLQMLTNSNRKQDPTNOQPDSCAPQONNG 120
QY 121 LPTPSADSGGGTTPATGGGGGDPSPATGGGGGDPPTATGGGGGDPPTATGGGGG 180
DB 121 LPTPSADSGGGTTPATGGGGGDPSPATGGGGGDPPTATGGGGGDPPTATGGGGG 180
QY 181 PRATGGGEGGVTPOITPOLANPNRTSGTSPDTAGSTEQAGKINVKPTIKYAGEVED 240
DB 181 PRATGGGEGGVTPOITPOLANPNRTSGTSPDTAGSTEQAGKINVKPTIKYAGEVED 240
QY 241 GGGATFTADKSMNGNDGEGNOKPMFELAGATLKNVNLGENEVDGIHVAKNAQEVYIDN 300
DB 241 GGGATFTADKSMNGNDGEGNOKPMFELAGATLKNVNLGENEVDGIHVAKNAQEVYIDN 300
QY 301 VHAQNVGEDLITVKGEGAAVTNLTKNSAKGADKVVQLNANTHLKIDNEFADDFGTA 360
DB 301 VHAQNVGEDLITVKGEGAAVTNLTKNSAKGADKVVQLNANTHLKIDNEFADDFGTA 360
QY 361 VRTNGKQFDDMSIELNGIEANHGKFAVKSDSDDLKLTATGNIAMTDVKNAYDKTOASTQ 420
DB 361 VRTNGKQFDDMSIELNGIEANHGKFAVKSDSDDLKLTATGNIAMTDVKNAYDKTOASTQ 420
QY 421 HTTEL 424
DB 421 HTTEL 424

RESULT 2
087327 PRELIMINARY: PRT: 441 AA.
AC 087327;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
DE HRP TYPE III SECRETED PROTEIN.
GN HRPW.
OS Pseudomonas syringae.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B728A;
RX MEDLINE-98422476; PubMed-9748456;
RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
  Collier A.;
RT "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
  to harpins and pectate lyases and can elicit the plant hypersensitive
  response and bind to pectate."
RL J. Bacteriol. 180:5211-5217(1998).
DR EMBL: AF037983; AAC62530.1; -.
SQ SEQUENCE 441 AA; 44714 MW; 7B3B3146E51897BB CRC64;

```

Query Match 72.3%; Score 1591.5; DB 2; Length 441;
 Best Local Similarity 70.7%; Pred. No. 6e-91;
 Matches 312; Conservative 42; Mismatches 70; Indels 17; Gaps 5;

```

QY 1 MSIGITPRP--QQTTPPLDPSALSGKSPQNTFGEONTQOALDPSALLFGSDTOKVNGF 58
DB 1 MSIGITPRPQQTTPPLDPSALSGKSPQNTFGEONTQOALDPSALLFGSDTOKVNGF 60
QY 59 PTPDSQONQDASRPNDQSNIAKLISALIMSLQMLTNSNRKQDPTNOQPDSCAPQONNG 118
DB 61 QPNTVQNPPTDSQATPDSQNVVLLSALVTSLQMLTNSNRKQDPTNOQPDSCAPQONNG 120

```

```

QY 119 GGLTPSADSGGGGTPPATGGG--GGDTPSATG--GGGGDPPTATGGG-----SGGGG 168
DB 121 GGLTPSADSGGGGTPPATGGG--GGDTPSATG--GGGGDPPTATGGG-----SGGGG 180
QY 169 TPTATGGGGGTP-----TATGGGEGGVTPOITPOLANPNRTSGTSPDTAGSTEQAGK 223
DB 181 SYVTGADSSGAPSTEDGTGGGGGSDGVTPOVTPOLANPNRTSGTSPDTAGSTEQAGK 240
QY 224 INVVKPTIKYAGEVEDGATFTADKSMNGNDGEGNOKPMFELAGATLKNVNLGENEY 283
DB 241 VNVVKPTIKYAGEVEDGATFTADKSMNGNDGEGNOKPMFELAGATLKNVNLGENEY 300
QY 284 DGIHVAKNAQEVYIDNVAQNVGEDLITVKGEGAAVTNLTKNSAKGADKVVQLNA 343
DB 301 DGIHVAKNAQEVYIDNVAQNVGEDLITVKGEGAAVTNLTKNSAKGADKVVQLNA 360
QY 344 NTHLKDNEKADDFGTVRTNGKQFDDMSIELNGIEANHGKFAVKSDSDDLKLTATGN 403
DB 361 DTHLKDNEKADDFGTVRTNGKQFDDMSIELNGIEANHGKFAVKSDSDDLKLTATGN 420
QY 404 AMTDVKNAYDKTOASTQHTTEL 424
DB 421 AMTDVKNAYDKTOASTQHTTEL 441

RESULT 3
091AM2 PRELIMINARY: PRT: 447 AA.
ID 091AM2;
AC 091AM2;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE HARPIN HRPW.
GN HRPW.
OS Erwinia amylovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwinia.
OX NCBI_TaxID=552;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EA246;
RA Kim J.F., Laby R.J., Beer S.V.;
RT "Comparison of the hrpN-flanking regions of two Erwinia amylovora
  strains with different host specificity."
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF083620; AAF63402.1; -.
SQ SEQUENCE 447 AA; 45340 MW; 0BBAEA3871EDC2F6 CRC64;

```

Query Match 25.6%; Score 563; DB 2; Length 447;
 Best Local Similarity 36.1%; Pred. No. 1.9e-27;
 Matches 147; Conservative 56; Mismatches 130; Indels 74; Gaps 13;

```

QY 72 KPNDQS--NIAKLISALIMSLQMLTNSNRKQDPTNOQPDSCAPQONNGSLG----- 122
DB 37 QPDRQPTTEOMAQLLELLKSL--LSPQSNATAGAGNDQTTGYNAGGLNGKRGITAG 93
QY 123 -TPSADS-----GGGG-----TPDAGGGG--GDP-----SATGGG 151
DB 94 TTPQSDSQNMLSEMGNNGLDQATITPPGGGGGQIGDNPFLKAMKLARMMDGSDGDFG 153
QY 152 GSDTPATGGGGGGGGTTPATGGG--SGGPTATGGGEGVTPOITPOL-----A 200
DB 154 GTGNSASASGTSSTSSGSPFNDLSGKAPSGNSPSSGVSPTSPTSPTSPLDPS 213
QY 201 NPKRTSG-----TGSVSDTAGS-----TEQAGKINVVDITIKVAGEVFDHGAT 245
DB 214 SPTKAKAGSTPTVDHPDVPVGSAGIGAGNSVATFSAGANOTVJLHDTTVKASQVFDKGT 273
QY 246 FTADKSMNGNDGEGNOKPMFELAGATLKNVNLGENEVDGIHVAKNAQEVYIDNVAQNV 305
DB 274 FTAGSELGDSGEGENOKPFILEDGASLKNVTVMGDDGADGSIHLXG-----DAKIDNLAHYTN 329

```



```

0Y      306 VEDILTYEGGCAVNTNNINKNSSAKKGGADVVOVNATHTKLIDNFKADDEGTWRTNG 365
       ||||| |||||   ::: | ||| : | : ||||| : | : | ||||| |||||
Db      330 VGEDAITYKPNSAGKKSHVEITNSSFEHNSDKLTLLQNMADTNLSVDNVKAKKDGETFVRING 389
       || : : : | : | ||| : ||||| : || : ||||| : || : ||||| : || : |||||
0Y      366 GKOFDMSIEIENGIEANHGKFALVKSDDSLKATGNIAMTDVKNAY 412
       | : | : | : | : | ||| : ||||| : || : ||||| : || : ||||| : || : |||||
Db      390 GQQ-CGMNDLNIHSIEDGRKFSVKSDESGLKYVNSDISSLVENHY 435
       | : | : | : | : | ||| : ||||| : || : ||||| : || : ||||| : || : |||||

RESULT      4
ID           054508      PRELIMINARY;          PRT;         447 AA.
AC           054508;
DT      01-JUN-1998 (TREMBLrel. 06. Created)
DT      01-NOV-1998 (TREMBLrel. 08. Last sequence update)
DT      01-NOV-1998 (TREMBLrel. 08. Last annotation update)
DE      HRPW PROTEIN.
GN      HRPW.
OS      Erwiniya amylovora.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Erwiniya.
OX      NCBI_TaxID=552;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CFBP1430;
RX      MEDLINE=9808611; PubMed=9426142;
RA      Gaudreault S., Malandrin L., Paulin J.P., Barry M.A.;
RT      "DspsA, an essential pathogenicity factor of Erwiniya amylovora showing
RT      homology with AvrE of Pseudomonas syringae, is secreted via the Hrp
RL      secretion pathway in a DspB-dependent way.";
RN      [2]
RP      SEQUENCE OF 138-447 FROM N.A.
RC      STRAIN=CFBP1430;
RA      Gaudreault S., Brisset M.N., Barry M.A.;
RL      Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
RN      [3]
RP      SEQUENCE OF 1-138 FROM N.A.
RC      STRAIN-EA321;
RX      MEDLINE=98115919; PubMed=9448330;
RA      Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,
RA      Conlin A.K., Collmer A., Beer S.V.;
RT      "Homology and functional similarity of an hrp-linked pathogenicity
RT      locus, dsrEF, of Erwiniya amylovora and the avirulence locus avrE of
RT      pseudomonas syringae pathovar tomato.";
RL      Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN-EA321, ATCC 49947;
RA      Kim J.F., Zumoff C.H., Beer S.V.;
RT      "HRPW, a new harpin of Erwiniya amylovora, is a member of a family of
RT      pectate lyases.";
RL      Phytopathology 87:0-0(1997).
RN      [5]
RP      SEQUENCE FROM N.A.
RC      STRAIN-EA321, ATCC 49947;
RX      MEDLINE=98422475; PubMed=9748455;
RA      Kim J.F., Beer S.V.;
RT      "HRPW of Erwiniya amylovora, a new harpin that contains a domain
RT      homologous to pectate lyases of a distinct class.";
RL      J. Bacteriol. 180:5203-5210(1998).
DR      EMBL; Y13831; CAAT74158.1; -
DR      EMBL; U97504; AAC04849.1; -
DR      EMBL; U94513; AAC62314.1; -
SEQUENCE     447 AA; 45347 MW; 7A744DL69D6D9CF3 CRC64;

Query Match      25.4%; Score 559; DB 2; Length 447;
Best Local Similarity 36.1%; Pred.No.3.4e-27;
Matches 147; Conservative 54; Mismatches 132; Indels 74; Gaps 13.

72 KPNSQS--NIAKLSALIMSLQWLTSNKKQDPINOEQPPDSQAFOFNNGGIG----- 122
||||| :||| : ||| | : | : | ||| | : | : | ||| | : ||| | : ||| |

```

```
Db      37 QPDRQTIQMQLLAELIKSL---LSPSGMAATGAGCNDGTGGNNGINGRKGTAG   93
Qy     123 -TPSADS-----GGCG-----TPTATGGCG-GDTP-----SATGGG    151
           ||| |
Db      94 TTPDSQSMLSEMGNNGIDOAITPDGGCGGOIGDNPLKAMLKLTARMMDGSDFGP    153
           ||| |
Qy     152 GGDPTPATGGSGGCGGTFATGGCG--SGGPTATGGEGGVTPQITPOL-----A    200
           ||| |
Db      154 GTGNNSASSGSTSSGGSPPNDSLSCGAPSGNSDCMYSPVSTFSPTSTPSPLDPFS    213
Qy     201 NPNRTSG-----TGVSDFTAGS----TEQAQINVKDTIRYAGEVFHGAF       245
           ||| |
Db      214 SPTKAAGSTPYTDHDPVPGASAGIGNSVATTSAQANTVLHDITTYAKGYFDGKGT    273
Qy     246 ETADKSMGNGDOGENOKPMFEALBEGATLKNVLGENEVDGIHYKAKNAOEVTDNYHAON    305
           ||| |
Db      274 FTAGSELGSGGSENOCKPIFLEDGASLNVTMGDDGAGCIHLYG---DAKDILHTVN    329
Qy     306 VEDDITTVYGEGAAVTNLTINKNSSAKGDKPVVOLNATHLIKINDFKADDFETWRTNG    365
           ||| |
Db      330 VEEDATTVYPNSAGKSKSHVEITNSSHEHSADKLQDLNADTNLSVDNVAKKDFETVRTNG    389
Qy     366 GFQFDMISIELNGIEANHGKFALVKSDDSDDLKLATINIAMIPTYKHAY         412
           ||| |
Db      390 GGQ-GMWDLNLISHSEDCKFFSYKSDSGLVNTSIDLGVENHY          435

RESULT        5
Q9RHMO              PRELIMINARY:                PRT:       224 AA.
AC      O9RHMO:
DI      01-MAY-2000 (TREMBRel. 13, Created)
DI      01-MAY-2000 (TREMBRel. 13, Last sequence update)
DI      01-MAY-2000 (TREMBRel. 13, Last annotation update)
OS      PECCATE LYASE.
OS      Bacillus sp.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group;
OC      Bacillus/Staphylococcus group; Bacillus.
OX      NCBI_TaxId=1409;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-KSM-P15;
RA      Hatada Y., Koike K., Saito K., Kobayashi T., Susumu I.;
RT      "Amino acid sequence and possible Catalytic residues of a novel
RL      alkaline pectate lyase from alkaliphilic Bacillus.";
RL      Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AB011839; JAA87892.1; -.
KW      Lyase.
SQ      SEQUENCE      224 AA: 23751 MW; 99D04821B09DE523 CRC64;

Query Match                      14.5%; Score 318; DB 2; Length 224;
Best Local Similarity 40.7%; Pred No.1.le-12;
Matches 74; Conservative 33; Mismatches 61; Indels 14; Gaps 5

Qy     226 VKDKTKVGAGEVPFGHGATFTAD-KSMGNGDOGENOKPMFEALBEGATLKNVLGENEVD    284
           ||| |
Db      31 VHETLRIVPAQTGFQDGQTYVANPTLDGSOAEENOKPIFLLEGASLKNVVYIAPAAD    90
Qy     285 GIHVAKAKNAOEVTIDNVHAONGEDLLITYKGEGGAAYTMLNTINKNSAKKADRVQOLAN    344
           ||| |
Db      91 GVHCYG---DOTTINVIMEDVGEPALTALKSSG---TVNISGGAAVAAYARVQIINA     141
Qy     345 TLTKIDNFADGFMGTVRTNGKOPDMSIELNGIEANRGKFALVKSDDDLKLATGNIA      404
           ||| |
Db      142 GIIINIRNPADDIGALVRONGSTTY-KVMANYENCNISRYKAILIRTDS---STIGRIIV    197
Qy     405 MT 406
Db      198 NT 199
```

09X622
ID 09X622 PRELIMINARY; PRT: 222 AA.
AC 09X622;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE PECCATE LYASE PRECURSOR (EC 4.2.2.2).
GN PECA.
OS Bacillus sp. Bp-23.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=89769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BP-23;
RA Soriano M., Blanco A., Diaz P., Pastor F.I.J.;
RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ237980; CAB40884.1; -
KW Signal; Lyase.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 222 PECCATE LYASE.
SQ SEQUENCE 222 AA; 23233 MW; 7F29F585791C9682 CRC64;

Query Match 14.3%; Score 315; DB 2; Length 222;
Best Local Similarity 38.7%; Pred. No. 1.7e-12;
Matches 77; Conservative 33; Mismatches 75; Indels 14; Gaps 5;

QY 209 GSVSDTAGSTEQAGKINVKYKDTIKVAGVFDGCHGATFTADKS-MGNGDQGENOKPFEEL 267
DB 12 GLVASIEGWPAAPAAAPTVNSTIVPKGTYYDQGRTEFVNPSTLDGSGQAEKOKPYEFL 71
QY 268 AEGATIKVNLGNEVDGIVKAKNAOEVTIDVHAONVGEEDLITVKGGAATVNLNIK 327
DB 72 EAGATIKVNLGNEVDGIVKAKNAOEVTIDVHAONVGEEDLITVKGGAATVNLNIK 122
QY 328 NSSAKADKRVQVLANVHLKIDNFKADDFGIVRTNGKQFDDMSIELNGIEANHGKFA 387
DB 123 GGAAYKAYKDFVQMNASGTFINIKFRADIDIGKLVKRGSTSY-AVNMTLNLSNISKDS 181
QY 388 LVKSDSDDLKATNGTAMT 406
DB 182 IMRTDS--SVSOGKITNT 197
RESULT 7
034310 PRELIMINARY; PRT: 221 AA.
AC 034310;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
GN TVPA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azeredo V., Bertsche M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.U., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Hatedch J., Harwood C.R., Hempt A.,
RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Iraya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakel S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Potworilik S., Prescott A.M.,
RA Presecan E., Pujić P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivoita C., Roche E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serior P., Shin B.S., Soldo B.,
RA Sorokin A., Taccanti E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandendol M., Vanlier F., Vassartoli A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT *The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis*;
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (Nov-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017113; AAC67291.1; -
DR EMBL; Z99121; CAB15500.1; -
SQ SEQUENCE 221 AA; 24281 MW; AB324700DE573236 CRC64;

Query Match 13.2%; Score 291.5; DB 2; Length 221;
Best Local Similarity 36.6%; Pred. No. 4.9e-11;
Matches 68; Conservative 32; Mismatches 71; Indels 15; Gaps 4;

QY 226 VVDITKVGAGEFVDGATFTADKSMGNGDQGENOKPFEELAEATIKVNLGNEVDG 285
DB 31 VVHETITVPKNTTYDQKGFVAGKELGDSQSENDDPVRVEDGATIKVYVAGAPRAG 90
QY 286 IHVKAKNAOEVTIDVHAONVGEEDLITVKGGAATVNLNIKSSAKGADKRVQVLANV 345
DB 91 VHTYXG---NVNIONVWKEVDGDAITVKKEG-----KVITDGSQAKSADKIFQINKAS 141
QY 346 HAKIDNFKADDFGIVRTNGKQFDDMSIELNGIEANHGKFAVKSDDKLKATNGIAM 405
DB 142 TFFVKKFTADNGKRFIRQLGSGTF-HVDVLIIDKCTITNMKEALFRDS-----KTSTVR 195
QY 406 TDYKHA 411
DB 196 TMTTIS 201
RESULT 8
09EX16 PRELIMINARY; PRT: 266 AA.
AC 09EX16;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
GN 2SCG38.03.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN 121
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN 131
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA MEDLINE:97000351; Pubmed-8843436;
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinasli H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL445303; CACC13062.1; -;
 KW Lyase.
 SQ SEQUENCE 266 AA; 27499 MW; 7687BFD93717913 CRC64;

Query Match 10.98; Score 239; DB 2; Length 266;
 Best Local Similarity 33.88; Pred. No. 1.1e-07;
 Matches 67; Conservative 25; Mismatches 74; Indels 32; Gaps 7;

QY 202 PNRTSGTGSVDTAETGAEQAKINVKTIKVGAEVDPDGHGATFTADKSMGNGDQGENO 261
 Db 43 PNGSEGVSTLSVSGTDYGMK-----RLXGTGD-----LGGGQDEDO 81
 QY 262 KPMFLAEGATLKNVNLGENEVDGIHVAKAKNAOEVTIDNVHAONGEDLITVKEGGA 321
 Db 82 GPILFLAPGALVKNIITIGAPADGVHCKG---SCTLQNVMEVDGEDAATFRSSSSNV 137
 QY 322 TNLTKNSAAGCADDXYQVQANATHLKIKNKADDFGIMVT--NGGQFPDMSTIELNGI 379
 Db 138 --YTVSGGGAADKDFQFQAGCTLNISGFAYKNFETFRSCNCSCTQY-RTITINNGI 194
 QY 380 EAN--HGKFALVKSDDSD 395
 Db 195 EVNMKGGRAGINNTNGD 212

RESULT 9
 ID 001368 PRELIMINARY; PRT; 3190 AA.
 AC 001368;
 DT 01-JUL-1997 (TEMBLrel. 04, Created)
 DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE CREB-BINDING PROTEIN HOMOLOG.
 GN NEJ OR CG15319.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCB1_TaxID=7227;
 RN 111
 RP SEQUENCE FROM N.A.
 RA MEDLINE:97263578; Pubmed-9109493;
 RA Aikimaru H., Chen Y., Dai P., Hou D.X., Nonaka M., Smolik S.M.,
 RA Armstrong S., Goodman R.H., Ishii S.;
 RT "Drosophila CBP is a co-activator of cubitus interruptus in hedgehog
 RT signalling.";
 RL Nature 386:735-738(1997).
 DR EMBL: U08570; AAB53050.1; -;
 DR HSSP: P03622; 21FO.
 DR FLYBase: FBgn0015624; nej.
 DR InterPro: IPR000197; -;
 DR InterPro: IPR000433; -;
 DR InterPro: IPR001487; -;
 DR InterPro: IPR003101; -;
 DR Pfam: PF00439; bromodomain; 1.
 DR Pfam: PF00569; zz; 1.
 DR Pfam: PF02135; zf-TAZ; 2.

DR Pfam: PF02172; KIX; 1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR PROSITE: PS00633; BROMODOMAIN_1; UNKNOWN_1.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 DR SMART: SM00291; znf-zz; 1.
 SQ SEQUENCE 3190 AA; 331879 MW; E53526F78BC055A8 CRC64;

Query Match 9.88; Score 216; DB 5; Length 3190;
 Best Local Similarity 25.18; Pred. No. 5.2e-05;
 Matches 105; Conservative 37; Mismatches 147; Indels 130; Gaps 21;

QY 14 TPDLPSALSGSPQPNFTGEON--TQQAIDPSALLFGSDTKQDNVFPDPTVONPDAS 71
 Db 1305 TPSTLEALNAGAGAPGTGGSANVTVSAPSPSPGLSNP---SIGTP---SNNNSS 1356
 QY 72 KPNDSOSNIATLISALIMSLQMLTNSNKKQDTQDQPDQAPFQNNGL--GTP---- 124
 Db 1357 SANNNPVSLSLMQ-----QPLSN---RPGTPPYIPASPVPATSAAGLAASSTPASAA 1406
 QY 125 ----SADSGGGGTPDATTGGGGGTPSATGGGG-----DTPA-----TGGGGSGGGCT 169
 Db 1407 ATCASGSGSGSSSGAVPAAGASSTSSSSAGSGTPLSVSTPATMATSSGGCGGGGN 1466
 QY 170 PTATGGGSGCTPPTA-----TGEGEGGVTPQIT----- 196
 Db 1467 ---AGGSGTTPASPAPLLMSGGTAGGCTGATTTTSTSSSRMMSSSSLSQMAALEAA 1523
 QY 197 -----POLANPNTSGTGSVDTAETGAEQAKINVK--DTIKVGADEVDPDGHGA 244
 Db 1524 ARDNDETTPSPSGENTNGSGSGN---AGCMASKGLDSIKQDDIK-----K 1568
 QY 245 TFLTKDSMGNDQGENOKPMFLAEGATLKNV-----LGENEVDGIHVAKNA 293
 Db 1569 EFMDDSCGKNNDSSQ---MDCSTGGGKNVNDGTSMIKETKEDGLG-EVKK-T 1622
 QY 294 QEVITIDNVHAONGEDLITVKEGGAAN---TNLTKNSAKGADDXYQVQANATHL 348
 Db 1623 EAMVDENAGSTAGEH---HGEGGGSGVGGGKDNINGAHDGATGAGANDIKRTETK 1677

RESULT 10
 ID 092U23 PRELIMINARY; PRT; 730 AA.
 AC 092U23;
 DT 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
 DE F5F19.6 PROTEIN.
 GN F5F19.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OC NCB1_TaxID=3702;
 RN 111
 RP SEQUENCE FROM N.A.
 RA Vysotskaia V.S., Schwartz J.R., Tortum M., Yu G., Lenz C., Liu S.,
 RA Li J., Kremetskaia I., Lutos J., Gonzalez A., Altafi H., Araujo R.,
 RA Buehler E., Conway A.B., Conn L., Dunn P., Hansen N., Hutzler L.,
 RA Kim C., Palm C.J., Rowley D., Shinn P., Walker M., Davis R.W.,
 RA Becker J.R., Federspiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F5F19 sequence.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC006216; AAD12691.1; -;
 DR HSSP: P18670; 1JAC.
 DR InterPro: IPR001064; -;
 DR InterPro: IPR001229; -;
 DR Pfam: PF01419; Jaccalin; 3.
 DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 SQ SEQUENCE 730 AA; 73939 MW; E4323AF93B1D95E CRC64;

QY 355 DFGTAVRNG---GKQFDMST-----ELNGIEANIKRALYKSDSDIKLAT 400
Db 154 NDYKAYRSGNCKNGTWNCKRSVHMEGTAVKGGELIGINTNYGDKA----- 200
QY 401 GNIAMTDVHAYDKTQ 416
Db 201 ----TYSNCCPKTQ 211

RESULT 15

021835 PRELIMINARY; PRT: 1226 AA.
ID 021835
AC 021835;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DE 01-MAR-2001 (TREMBlrel. 16, last annotation update)
DE R08B4.1 PROTEIN.
GN R08B4.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Woldman P.,
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.*;
RL Nature 368:32-38(1994).
DR EMBL: Z68008: CAA92000.1; -.
DR InterPro: IPR001767; -.
DR InterPro: IPR001990; -.
DR InterPro: IPR003586; -.
DR Pfam: PF01079; Hint: 1.
DR PROSITE: PS00422; GRANINS_1; UNKNOWN_1.
DR SMART: SM00305; Hmtc; 1.
SQ SEQUENCE 1226 AA; 132756 MW; C41BF9A943884681 CRC64;

Query Match 8.9%; Score 196.5; DB 5; Length 1226;
Best Local Similarity 26.9%; Pred. No. 0.00027;
Matches 94; Conservative 24; Mismatches 156; Indels 75; Gaps 14;

QY 3 IGTFRP-----QQTTPDLDSALSG-----KSPQNTFGE 33
Db 710 IGRTPYPHONVYNDISAKFPNCPADLSGLSGMACCDGGIQLFEANKYIDQAKQEPDFDK 769
QY 34 QNTQQAIDPSALFLGSDTQKDVNFGTPTVQNPQDAS-KPNDQSNIKLISALIMSL 92
Db 770 HNTNRL-----AKIMTRAVOK--RCGTTFESVVALEADPSWGNKFNKRTCKIDS----- 816
QY 93 QMLTNSNKKQDTNQEQPDQS-APFQNNGLGTPSADSGGG-----TPDANGGGGDTPSA 147
Db 817 -QGYNALTYQSSSKPPPSDFIDIPNDPTLGGPTGSSGGCGGCGGSGGGGGSG 875
QY 148 TGGGGGDTPTATGGGGSGGGGTPTATGGSGGTPTATGGGGGVTPTPTPOLANPRTSG 207
Db 876 GSGGGGNSNSGGGGGGGG-----NGGGGNGNGGAGAGNGG-----AGANGNG 922

QY 208 TGSVSDTAGSTBOAGKINVVKDTIKVAGEFVFGHGAFTTADKSMGND-QGENOKPMFE 266
Db 923 AG-----AGNGAGAGAGN-----GNGAGAG--NGNGAGAGNGNGAGAGDASAAAAAQO 970
QY 267 LAEGATLKRVNLTGENEVDGSIHVAKNAQEVTTIDNVHAQNVGEDLTIVKG 315
Db 971 AAAAAQAQAAAAAAQAQAAAAAAQAAAAAAANPALSALVAATG 1019

Search completed: October 4, 2001, 22:33:09
Job time: 395 sec

Fri Oct 5 10:04:33 2001

us-09-597-513-2.rspt

Page 9

THIS PAGE BLANK (USPTO)

QY	61	ctgagcgcgagatttcgttgtatagaaggtgtgtactgtgtcaatttgtcaatttcaaaag	120
Db	74	ctgagcgccgagatttgcgttgatmaagggtgtgtactgtgtcaatttgtgtcattttcaag	133
QY	121	ccctctatgtgcgtgtgagagcaataccagctcttcctcgtcgtgtgtgtcaactgtgtgc	180
Db	134	ccctctatgtgcgtgtgagagcaataccagctcttcctcgtcgtgtgtgtcaactgtgtgc	193
QY	181	aggcatagcatttcagttcctgtcgttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	240
Db	194	aggcatagcatttcagttcctgtcgttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	253
QY	241	acagtgcaatgagatgcgcggcaaaacggggaaccggtgcgtgcgtgtgtgtgtgtgtgtgt	300
Db	254	acagtgcaatgagatgcgcggcaaaacggggaaccggtgcgtgcgtgtgtgtgtgtgtgtgt	313
QY	301	agcaagctaaccccccaaacatccatccctatcgaaacgagacgatacgcgcaacttgc	360
Db	314	agcaagctaaccccccaaacatccatccctatcgaaacgagacgatacgcgcaacttgc	373
QY	361	tctgttaaaccttggagctgtgcgttcggttccaatgtcccactgtgagatgaacgcagat	420
Db	374	tctgtttaaacccttggagctgtgcgttcggttccaatgtcccactgtgagatgaacgcagat	433
QY	421	gagcatcggaatcaacaccccgccgcgcgaacacgacacacgacgcaactgtatttcgcgcgt	480
Db	434	gagcatcggaatcaacaccccgccgcgcgaacacgacacacgacgcaactgtatttcgcgcgt	493
QY	481	aagcggaagaggtgtcttaaccaacacagttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	540
Db	494	aagcggaagaggtgtcttaaccaacacagttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	553
QY	541	cccgatgcacatgttgttcgc	600
Db	554	cccgatgcacatgttgttcgc	613
QY	601	cagcagcttccagaatccgc	660
Db	614	cagcagcttccagaatccgc	673
QY	661	taattgtatcagtgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgat	697
Db	674	taattgtatcagtgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgat	710
RESULT	6		
LOCUS	U94513	2708 bp	DNA
DEFINITION	U94513	2708 bp	DNA
ACCESSION	U94513	2708 bp	DNA
VERSION	U94513.1	GI:3414585	
KEYWORDS	type III protein secretion; Hrp pathway; hypersensitive response; pectate lyase.		
SOURCE	Erwinia amylovora.		
ORGANISM	Erwinia amylovora.		
REFERENCE	1 (bases 912 to 2255)		
AUTHORS	Kim,J.F., Zumoff,C.H. and Beer,S.V.		
TITLE	HrpW, a new harpin of Erwinia amylovora, is a member of a family of pectate lyases		
JOURNAL	Phytopathology 87, 552 (1997)		
REFERENCE	2 (bases 1 to 2708)		
AUTHORS	Kim,J.F. and Beer,S.V.		
TITLE	HrpW of Erwinia amylovora, a new harpin that contains a domain homologous to pectate lyases of a distinct class		
JOURNAL	J. Bacteriol. 180 (19), 5203-5210 (1998)		
MEDLINE	98422475		
REFERENCE	3 (bases 1 to 2708)		
AUTHORS	Kim,J.F. and Beer,S.V.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-MAR-1997) Plant Pathology, Cornell University, 334		

THIS PAGE BLANK (USPTO)